

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd

OM nucleic - nucleic search, using sw model

Run on: July 27, 2004, 09:14:42 ; Search time 2079 Seconds  
(without alignments)  
5229.234 Million cell updates/sec

```
Title: US-09-714-865B-1
Perfect score: 2224
Sequence: 1 acttgagtcacatggggg.....agctctgtgttttgatgca 2224
```

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3216467 seqs, 2444149694 residues

Total number of hits satisfying chosen parameters: 6432934

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

```
Database : Published Applications NA.*
1: /cgm2_6/ptcdatata/1/pub/pna/US07_PUBCOMB.seq.*
2: /cgm2_6/ptcdatata/1/pub/pna/PCR_NEW_PUB.seq.*
3: /cgm2_6/ptcdatata/1/pub/pna/US05_NEW_PUB.seq.*
4: /cgm2_6/ptcdatata/1/pub/pna/US05_PUBCOMB.seq.*
5: /cgm2_6/ptcdatata/1/pub/pna/US07_NEW_PUB.seq.*
6: /cgm2_6/ptcdatata/1/pub/pna/PCRUS_PUBCOMB.seq.*
7: /cgm2_6/ptcdatata/1/pub/pna/US08_NEW_PUB.seq.*
8: /cgm2_6/ptcdatata/1/pub/pna/US08_PUBCOMB.seq.*
9: /cgm2_6/ptcdatata/1/pub/pna/US09_PUBCOMB.seq.*
10: /cgm2_6/ptcdatata/1/pub/pna/US09_PUBCOMB.seq.*
11: /cgm2_6/ptcdatata/1/pub/pna/US09C_PUBCOMB.seq.*
12: /cgm2_6/ptcdatata/1/pub/pna/US09_NEW_PUB.seq.*
13: /cgm2_6/ptcdatata/1/pub/pna/US09_NEW_PUB.seq.*
14: /cgm2_6/ptcdatata/1/pub/pna/US10_PUBCOMB.seq.*
15: /cgm2_6/ptcdatata/1/pub/pna/US10B_PUBCOMB.seq.*
16: /cgm2_6/ptcdatata/1/pub/pna/US10C_PUBCOMB.seq.*
17: /cgm2_6/ptcdatata/1/pub/pna/US10_NEW_PUB.seq.*
18: /cgm2_6/ptcdatata/1/pub/pna/US60_NEW_PUB.seq.*
19: /cgm2_6/ptcdatata/1/pub/pna/US60_PUBCOMB.seq.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1.601	72.0	1.984	16	US-10-103-047-1652	Sequence 1652, Ap
2	389.8	17.5	420	9	US-09-633-381-1475	Sequence 1475, Ap
3	386.4	17.4	2451	14	US-10-004-090-480	Sequence 480, Ap
4	384.8	17.3	4423	14	US-10-002-600-67	Sequence 67, Ap
5	383.2	17.2	4416	15	US-10-293-862-4	Sequence 4, Ap
6	356	16.0	591	15	US-09-766-891-1170	Sequence 1170, Ap
7	327	14.7	2019	15	US-10-032-585-6564	Sequence 6564, Ap
8	301.6	13.6	2362	17	US-10-437-963-58448	Sequence 58448, Ap
9	298.2	13.4	1815	16	US-10-369-493-46077	Sequence 46077, A
10	298	13.4	1854	16	US-10-369-493-46167	Sequence 46167, A
11	293.4	13.2	1908	16	US-10-369-494-46378	Sequence 46378, A
12	282.2	12.7	3012	13	US-10-424-589-14656	Sequence 14656, A
13	282.2	12.4	2672	13	US-10-424-593-14660	Sequence 14660, A
14	276.2	12.4	1681	16	US-10-369-493-26954	Sequence 26954, A

15	275.8	12.4	1902	9	US-09-938-8424-521	Sequence 521, App
16	275.8	12.4	1902	11	US-09-938-8424-5217	Sequence 521, App
c 17	273	12.3	533	15	US-10-029-386-12207	Sequence 12207, A
c 18	272.8	12.3	1705	16	US-10-260-238-490	Sequence 490, App
c 19	271	12.2	271	15	US-10-029-386-23597	Sequence 25907, A
20	268.4	12.1	2025	15	US-10-128-714-2551	Sequence 2551, Ap
21	268.4	12.1	2025	15	US-10-128-714-7551	Sequence 7551, Ap
22	268.4	12.1	2105	15	US-10-128-714-1551	Sequence 1551, Ap
23	268.4	12.1	2105	15	US-10-128-714-6551	Sequence 6551, Ap
24	268.4	12.1	4105	15	US-10-128-714-551	Sequence 551, App
25	268.4	12.1	4105	15	US-10-128-714-5551	Sequence 5551, Ap
26	268.4	12.1	1616	13	US-10-424-599-9376	Sequence 9376, Ap
27	229.8	10.3	2519	17	US-10-437-963-60284	Sequence 60284, A
28	220	9.9	2021	16	US-10-320-797-1250	Sequence 1250, A
29	220	9.9	2194	10	US-09-764-891-7481	Sequence 7481, Ap
30	220	9.9	2195	10	US-09-764-891-7482	Sequence 7482, Ap
31	220	9.9	2197	10	US-09-764-891-7483	Sequence 7483, Ap
32	220	9.9	4021	16	US-10-320-797-250	Sequence 250, App
33	219.2	9.9	1860	16	US-10-320-797-2250	Sequence 2250, Ap
34	210.2	9.5	3841	15	US-10-198-846-12716	Sequence 12716, A
35	206	9.3	2619	17	US-10-437-963-46500	Sequence 46500, A
36	204.8	9.2	1884	17	US-10-437-963-19215	Sequence 19215, A
37	191.4	8.6	1606	13	US-10-424-599-21465	Sequence 21465, A
38	188.8	8.5	2596	13	US-10-425-114-22685	Sequence 22685, A
39	178.2	8.0	2236	13	US-10-424-599-55854	Sequence 55854, A
40	175.2	7.9	1962	13	US-10-425-114-31433	Sequence 31433, A
41	173	7.8	2934	17	US-10-437-963-69859	Sequence 69859, A
42	172	7.7	1538	13	US-10-425-114-23650	Sequence 23650, A
43	172	7.7	1904	13	US-10-469-493-28786	Sequence 28786, A
44	170.2	7.7	1176	16	US-10-369-493-14361	Sequence 14361, A
45	168.4	7.6	403	9	US-09-960-352-14361	Sequence 14361, A

ALIGNMENTS

RESULT 1

US-10-104-047-1652

Sequence 1652, Application US/10104047

Publication No. US20030236392A1

GENERAL INFORMATION:

APPLICANT: HELIX RESEARCH INSTITUTE

TITLE OF INVENTION: No. US20030236392A1el full length cDNA

FILE REFERENCE: H1-A0105

CURRENT APPLICATION NUMBER: US/10/104,047

CURRENT FILING DATE: 2002-03-25

PRIOR APPLICATION NUMBER:

NUMBER OF SEQ ID NOS: 4096

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 1652

LENGTH: 1984

TYPE: DNA

ORGANISM: Homo sapiens

US-10-104-047-1652

Query Match

Best Local Similarity 96.9%; Pred No. 0;

Matches 1662; Conservative

72.0%; Score 1601; DB 16; Length 1984;

Indels 48; Gaps 1

510 ATATGACTTAGACCCAGACGAGATATGACAGCGACCTGCTGGCTTTTGGTCTTAGAA 569

117 ATATGACTTAGACCCAGACGAGATATGACAGCGACCTGCTGGCTTTTGGTCTTAGAA 176

570 GACCGAGTATTAAGTGCGACAGGTAATGGTATACCTTCTAAAGCAGAACTGGCAGTGGAA 629

177 GACCGAGTATTAAGTGCGACAGGTAATGGTATACCTTCTAAAGCAGAACTGGCAGTGGAA 236

630 GTGAACGAGGTGGTTCACAAAGGTTTAAATGAAGAAGTAATACAGGCTCTGGAAGAAATT 689

237 GTGAACGAG-----ATT 248

690 CTTGAAGTCAAGACAGAGAGAGAAAGTATGTCATCTCAAGGACCAAAAGTACCT 749

Db	249	CTTGGAAAGTCAGAAAGAGAGAGAGAAAGTATGATCTCGAGGACCCAAAAGTGA	308
Qy	750	ACATACCCCTCTCCACCTGAGATGAGACTCCATCTTTGCATTAATCGACAGCA	809
Db	309	ACATACCCCTCTCTCCACCTGAGAGTAGAGCTCCATCTTTGCATTAATCAGACGA	368
Qy	810	TAAACTGCAAAATACGACATATCTTGTGTGGAAGTGTCTGACATATGACACAGC	869
Db	369	TAACTTGCACAAATTCGACACTATTTCTTGGAAGTGTCTGACATATGACACAGC	428
Qy	870	CAATTCGTACTTTGGAAGGCTAATCTGTGACACATGAAATAACAATTGTCTAAG	929
Db	429	CAATTCGTACTTTGGAAGGCTAATCTGTGACACATGAAATAACAATTGTCTAAG	488
Qy	930	CTGGTTATATCTAAGTTACTCTGTGCAAAAATAAGATTTCTTATCTTGTGAGAC	989
Db	489	CTGGTTATATCTAAGTTACTCTGTGCAAAAATAAGATTTCTTATCTTGTGAGAC	548
Qy	990	GAGATTTGATGGCTTGCTCAACAGGGTCTGGGAAGACTGGCTTTTCTCTACAA	1049
Db	549	GAGATTTGATGGCTTGCTCAACAGGGTCTGGGAAGACTGGCTTTTCTCTACAA	608
Qy	1050	TTTGGCTCATATGATGCAATGGAATATCTGCCAGTGTCTTAAAGATTGCAAGAC	1109
Db	609	TTTGGCTCATATGATGCAATGGAATATCTGCCAGTGTCTTAAAGATTGCAAGAC	668
Qy	1110	CAGAGTATTTATTTGAGACCACTCGAGAAATTTGATCCAGATTAATTTGGAAGCA	1169
Db	669	CAGAGTATTTATTTGAGACCACTCGAGAAATTTGATCCAGATTAATTTGGAAGCA	728
Qy	1170	GAAATTTTCTTTGGGACTTGTGTAAAGCTGTGTTATATGAGGGGAAACCAAGCTGG	1229
Db	729	GAAATTTTCTTTGGGACTTGTGTAAAGCTTGTATATATGAGGGGAAACCAAGCTGG	788
Qy	1230	GACATTCGAATTCGAAATATGAGACAGCTGTAAATATTTATGCTCTCTCGGAAGAC	1289
Db	789	GACATTCGAATTCGAAATATGAGACAGCTGTAAATATATGCTCTCTCGGAAGAC	848
Qy	1290	TGATGATATCATAGCGAAAGAAAGATTTGTCTCAACAGATCAAAATCTAGTTTGG	1349
Db	849	TGATGATATCATAGCGAAAGAAAGATTTGTCTCAACAGATCAAAATCTAGTTTGG	908
Qy	1350	ATGAAGCTGATTCGCATTTTGTGATATGGGTTTGTGCCGAATATGAAGATTAATTTCT	1409
Db	909	ATGAAGCTGATTCGCATTTTGTGATATGGGTTTGTGCCGAATATGAAGATTAATTTCT	968
Qy	1410	GCCGAGATCCCATCAAGAGACAGCGCCAAACCCCTATGTCTAGTCACACTTTCCAG	1469
Db	969	GCCGAGATCCCATCAAGAGACAGCGCCAAACCCCTATGTCTAGTCACACTTTCCAG	1028
Qy	1470	AGGAATTCAAAGTTGGCTCGAAGATTTTAAAGTCAAAATATCTGTGTCTGTG	1529
Db	1029	AGGAATTCAAAGTTGGCTCGAAGATTTTAAAGTCAAAATATCTGTGTCTGTG	1088
Qy	1530	GACAAGTCGGGAGCGCATGTAGAGATGTTCAAGCAACCGTCTCGAAGTGGCCAGTCT	1589
Db	1089	GACAAGTCGGGAGCGCATGTAGAGATGTTCAAGCAACCGTCTCGAAGTGGCCAGTCT	1148
Qy	1590	CAAAAAGAAAGCTCGTTGAAAATTCGCGAAAATATGGGATGAAAGAACTATGTCT	1649
Db	1149	CAAAAAGAAAGCTCGTTGAAAATTCGCGAAAATATGGGATGAAAGAACTATGTCT	1208
Qy	1650	TTGTGTGAATCTAAAGAAAAGACAAATTTTATCTGCACTTTCTTGTGCAAAAAATAT	1709
Db	1209	TTGTGTGAATCTAAAGAAAAGACAAATTTTATCTGCACTTTCTTGTGCAAAAAATAT	1268
Qy	1710	CAACTACAGATCATGTGATGCGGAAACAGAGACGGGAGCAAGCTTTGTGAGATT	1769
Db	1269	CAACTACAGATCATGTGATGCGGAAACAGAGACGGGAGCAAGCTTTGTGAGATT	1328
Qy	1770	TTGCTTTGAAAAGTCCCAATTTTGTGTCTACTTCAATGACTGCGCAAGAGGCTGAGTA	1829

Db	1329	TTTCGCTTGGAAAGCGCCACAGTCTCTGTGTTCGTAATCTCAATGACTGCAGAGGCGTGGAATA	1388
Oy	1830	TTGAAAATGTGCACAATGTTATCAATTATTTGATCTTCTCTTCAACATTTGATGAATATGTTTC	1888
Db	1389	TGGAATAATGTGCACATGTATTCATATTTGATTCCTCTCTTCAACATGATGAATATGTTTC	1448
Oy	1890	ATCGAATTTGGGGGTCTCGTGTGGTGTGGGAATACAGCGACAGAGCATATTCCTTTTTGATGC	1948
Db	1449	ATCGAATTTGGGGGTCTCGTGTGGTGTGGGAATACAGCGACAGAGCATATTCCTTTTTGATGC	1508
Oy	1950	TTGAAATCGGAATPACCATTTTAGACACAGCCTCTAGTAAAAAGTATTGACAGATGCTCAACAGG	2008
Db	1509	TTGAATCGGAATPACCATTTTAGACACAGCCTCTAGTAAAAAGTATTGACAGATGCTCAACAGG	1568
Oy	2010	ATGTTCTCTGCATGTTGTGGAAAGAAATATGTCCTTTTAGAACATAATTCCTTGGCTTCACTGGTA	2068
Db	1569	ATGTTCTCTGCATGTTGTGGAAAGAAATATGTCCTTTAGAACATAATTCCTTGGCTTCACTGGTA	1628
Oy	2070	GTACAGAAGAAACGTGTTTGCATCAGTTGATATACAGAAAGGCGACAGACATTTGAAACA	2128
Db	1629	GTACAGAAGAAACGTGTTTGCATCAGTTGATATACAGAAAGGCGACAGACATTTGAAACA	1688
Oy	2130	CAGCTGGGTTTTCTTCTTACAGAGCTCCCAAATCCAGTATGATGATGAGTCTATGGGATTTAA	2189
Db	1689	CAGCTGGGTTTTCTTCTTACAGAGCTCCCAAATCCAGTATGATGATGAGTCTATGGGATTTAA	1748
Oy	2190	GCCAAAACATCCTTCAAGTCTGTGGTTTTGATGCA	2224
Db	1749	GCCAAAACATCCTTCAAGTCTGTGGTTTTGATGCA	1783
 RESULT 2 US-09-833-381-1475 ; Sequence 1475, Application US/09833381 ; Patent No. US20020132090A1 ; GENERAL INFORMATION: ; APPLICANT: Robison, Keith E. ; TITLE OF INVENTION: No. US20020132090A1el Nucleic Acid and Protein Homologs ; FILE REFERENCE: 5800-119 ; CURRENT APPLICATION NUMBER: US/09/833,381 ; CURRENT FILING DATE: 2001-04-11 ; PRIOR APPLICATION NUMBER: 09/516,448 ; PRIOR FILING DATE: 2000-02-29 ; NUMBER OF SEQ ID NOS: 2050 ; SOFTWARE: FastSeq for Windows Version 3.0 ; SEQ ID NO 1475 ; LENGTH: 420 ; TYPE: DNA ; ORGANISM: Homo sapiens US-09-833-381-1475			
 Query Match                    17.5%; Score 389.8; DB 9; Length 420; Best Local Similarity         99.5%; Pred. No. 3.1e-100; Matches 391; Conservative    0; Mismatches 2; Indels 0; Gaps 0;			
Oy	618	GACAAATACCACTATTTCTTGTGTGGAGTGTGTGGAAATGATGACACACACGCAATTTCTG	877
Db	28	GTCAAAATACCACTATTTCTTGTGGAAGTGTGTGGACATATGACACACGCAATTTCTG	87
Oy	878	ACTTTTGAAGAGCTAATCTCTGTGCAGACATGAAATACAACTTGTAAAGCTGTTAT	937
Db	88	ACTTTTGAAGAGCTAATCTCTGTGCAGACATGAAATACAACTTGTAAAGCTGTTAT	147
Oy	938	ACTAAGCTTACTCTGTGTGCAAAAATACAGTATTTCTTACTACTTGTGACGAGACGAGATTGG	997
Db	148	ACTAAGCTTACTCTGTGTGCAAAAATACAGTATTTCTTACTACTTGTGACGAGACGAGATTGG	207
Oy	998	ATGACTTGCTGTCAAAACAGGAGTGTGGGAAGACAGCGGCTTTTCTCCACCAATTTTGACT	1057
Db	208	ATGACTTGCTGTCAAAACAGGAGTGTGGGAAGACAGCGGCTTTTCTCCACCAATTTTGACT	267
Oy	1058	CATATGATGATGATGGAATPACTGCCAGTCTGTTTTAAAGATTGACGAAACGAGAGTGT	1117

Db		268	CATVATGATGATGGAAATTAACGTCCAGTCGCTTTTAAAGATGTGCAGGAACAAGATGT	327
Oy		1118	ATTATTTGATACCACCAACTCGAGAATTGGTGCAACCAATTTATTTGGAACCCGAAAATTT	1177
Db		328	ATTATTTGATCACCACCACTCGAATTTGGTGCAACCAATTTATTTGGAACCCGAAAATTT	387
Oy		1178	TCTTTGGAGCTTGTGTGAAGCTGTGTATA	1210
Db		388	TCTTTGGAGCTTGTGTGAAGCTGTGTATA	420
 RESULT 3 US-10-044-090-480 Sequence 480, Application US/10044090 Publication No. US20020137081A1 GENERAL INFORMATION: APPLICANT: Olga Bandman TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION FILE REFERENCE: PA-0028 US CURRENT APPLICATION NUMBER: US/10/044,090 CURRENT FILING DATE: 2002-01-09 NUMBER OF SEQ ID NOS: 850 SOFTWARE: PERL Program SEQ ID NO 480 LENGTH: 2451 TYPE: DNA ORGANISM: Homo sapiens FEATURE: NAME/KEY: misc feature OTHER INFORMATION: Incyte ID No. US20020137081A1 035282CB1 US-10-044-090-480				
Query Match		17.4%	Score 386.4; DB 14; Length 2451;	
Best Local Similarity		58.7%	Pred. No. 9,6e-99;	
Matches 763; Conservative		0;	Mismatches 501; Indels 36; Gaps 4;	
Oy		801	AGAGAGCATTAATTTGGAAATATGACACATTTCTTGGAAAGTGTGACATGATG	860
Db		503	ACACGGGAGTTAATTGAGAAATATGATATATACAGTAGAGGCCAACCGAGTAACT	562
Oy		861	CACCAACAGCAATTTCTACTTTTGAAGAAGCTATCTGTGCAGACATGAATTAACA	920
Db		563	GTCCTCCACAAATTTGAAGATTTTAGGATATGTACATGGAGAAATTAATCATGGGGA	622
Oy		921	TTGCTAAAGCTGGTTATTAAGTTATCTCTGTGCAAAATATCAATTTCTATATAC	980
Db		623	TTGAAGTACTCGGTATTAAGTGTCTCACTCCAGATGCAAAACATGCGCATTCATATATA	682
Oy		981	TTGAGAGCAGAGATTTATGATGGCTGTGCTCAACAGGAGTCTGGAAGACTGGCGCTTTTC	1040
Db		683	AGGAAAAAAGAGCTTAATGGCTTGTGCCCAAACAGATCTGGGAAAACTGCAGATTTC	742
Oy		1041	TCTCAACAATTTGGCTCA-----TATAGCATGATG	1073
Db		743	TTTTACCAATTAAGTCAAGTATATACAGATGATGCAGAGAACTTTGAAGGCTGTGA	802
Oy		1074	GAATTAATCGCACGCTTTTAAAGATGTGAGGAACAGAGTATATTTAGACACAA	1133
Db		803	AGGTAATTTGAAGGTATGGCGCGCGCAACAAATATCCAATATCTTGTTTAGCCCCAA	862
Oy		1134	CTCGAGAATTTGGTCAACCAAGATTTATTTGGAAAGCCGAAATTTCTTTGGAGCTGTG	1193
Db		863	CAGAGAAATTTGGCTGTGACAGTCTATAGAGAGCCGAAATTTCTTCACCGATCTAGAG	922
Oy		1194	TAAAGCTGTGTATATATAGGGGGAACCAAGCTGGGACATTGAATTCACAAATGTAC	1253
Db		923	TTCTGCTGTGTGATTTATGTGTGTGTCTGATTTGTGTGCGCAGATTCGGGACTTGAAC	982
Oy		1254	AAGGCTGTATATATTTATGTGCTACTCTCTGGAAAGCTGATGATATCATAGCCAAAGAA	1313
Db		983	GTGATGCACTTGTATGTATGCCACTCCAGAGAGCTTATGATGATATGAGAAAGGAA	1042
Oy		1314	AGATTTGGCTCAACACAGATCAAAATTAATTAATTTGGATGAAGCTGATCGATTTGGATA	1373

Db	1043	AGATTGATTAGACTTCTGCAAGTACTAGTGTGGATGAAGCTGATAGATGCTGGATA	1102
Qy	1374	TGGGTTTTGGTCCAGAAATGAAGAAGTAAATTTCTTGCCCAAGATGCGATCAAGAAGAC	1433
Db	1103	TGGGATTTGGAACCTCCAGATACGTCGTATAGTTGAACAAGATACTATGCGACCAAAAGGCG	1162
Qy	1434	AGCGCCAAACCCCTATGTCAGTGCACATTTTCCAGAGAAATTCAAAGCTGGCTGCAG	1493
Db	1163	TTGCTCACAACATCATGTTTAGTGTCTACTTTTCTTAAGAAATACAGATGCTGTGCTG	1222
Qy	1494	AGTTTTTAAAGTCAAAATTAATCTGTTTGTGTGCTTGGCAAGGTGGGTGAGCATGTAGAG	1553
Db	1223	ACTTTTGGATGAA---TATATCTTTTGGCTGTAGGAGAGTGGGCTCTACCTCTGAGA	1279
Qy	1554	ATGTTCAAGCAGACCGGTTCTCCAAATTTGGCCGTTCTCAAAAAGAAAGCTCGTTGAA	1611
Db	1280	ACATCAACACAGAAAGTGTGTTGGGTGGAAGCTTAGATMAAGGCTCATTTTCTACTGACA	1333
Qy	1614	TTCTGCAAAACATAGG---GATGAAAGACATATGTTCTTTGTGAAACTAAGAAAAAG	1670
Db	1340	TTTTAGGTGCAACAGGAGATATCACTTACTTATGTTGTGTGAGACCAAAAAGGAG	1399
Qy	1671	CAGATTTTACTGCAACTTTTCTTGTGCAAGAAAAATATCAACTCAAGATTCATGGTG	1730
Db	1400	CAGATTCCTGGAGGATTTCTTATACATGAAGATATGCTTGTACTATGATATTCATGGAG	1455
Qy	1731	ATCGGAAACAGAGAGAGCGGAGCAGCTCTTGAGATTTTCGCTTTGGAAAGTCCCGAG	1790
Db	1460	ACCGGTACAGAGAGATCGAAGAGAGGCCCTTCAACGATTTGCTCAGGAAAAAGCCCA	1515
Qy	1791	TTCTTTGTGCTACTTTCAGTACTGCTCCAGAGGCTGAGATTTGAATATGTGAACTGTTA	1851
Db	1520	TTCTAGTGGCTACAGCTGCTGGCACAGAGACATGACATTTCAAAATGTGAGACATGTTA	1575
Qy	1851	TCATTTTGTGATCTTCCTCTCTCAATTTGATGAATATGATCATGCAATTTGGGAGTACTGCTC	1910
Db	1580	TCAATTTTGAATTTGCGCAAGTATATGAAGAAATATGTGCATCGTATTTGGCGGTACAGAC	1633
Qy	1911	GTTGTGGAAATACTGGCAGACAAATTTCTTTTGTATCTTGAATCGATTAACATTTAG	1975
Db	1640	GTTAGAGAAACCTGGGCTTGGCCACCTCATTTCTTTAA---TGAAAAAAATATGAATATTA	1690
Qy	1971	CACAGCCCTACAGTAAAGTATTTGACAGATGCTCAACAGAGATGTTCCGATCGTGGAG	2033
Db	1697	CAAAGGATTTTGTGATTTCTTTGTGAAGCTTAACAGAACTGCTTCTTGTTGGAAA	1755
Qy	2031	AAATGCTTTAGTACATACATTCCTGGCTTCAGTGGTAG	2070
Db	1757	ATATGGCTTATGAACACCATCAAAAGGTGGCAGCGTGG	1796
RESULT 4			
US-10-002-600-67			
Sequence 67, Application US/10002600			
Publication No. US20020137077A1			
GENERAL INFORMATION:			
APPLICANT: Hopkins, Christopher M.			
APPLICANT: Peterson, David P.			
APPLICANT: Cocks, Benjamin G.			
APPLICANT: Hawkins, Phillip R.			
TITLE OF INVENTION: GENES REGULATED IN ACTIVATED T CELLS			
FILE REFERENCE: PA-0042 US			
CURRENT APPLICATION NUMBER: US/10/002,600			
CURRENT FILING DATE: 2001-10-25			
PRIOR APPLICATION NUMBER: 60/243,521			
PRIOR FILING DATE: 2000-10-25			
NUMBER OF SEQ ID NOS: 116			
SOFTWARE: PERL Program			
SEQ ID NO 67			
LENGTH: 4423			
TYPE: DNA			
ORGANISM: Homo sapiens			

FEATURE:  
; NAME/KEY: misc.feature  
; OTHER INFORMATION: Template ID: 350667.1  
US-10-002-600-67

Query Match 17.3%; Score 384.8; DB 14; Length 4423;  
Best Local Similarity 58.6%; Pred. No. 4.1e-98;  
Matches 762; Conservative 0; Mismatches 502; Indels 36; Gaps 4;

QY 801 AGACAGCATTAAGTTCGACAAATAGACACTATTCCTGTGGAAAGTGTGGACATGATG 860  
DB 536 ACACGGGGATTTAACTTTGAGAAATATGATATACAGTAGAGGCAACCGGAGTAATCT 595  
QY 861 CACACAGCAATTCCTGACTTTTGAAGAACTAATCTGTGACAGACTGAATPACACA 920  
DB 596 GTCTCCACATATGAAATTTTAGCATATGACATGGGAAATTTATCATGAGGAGACA 655  
QY 921 TTGTAAAGCTGTTATACCTTAAGCTTAATCTGTGCAAAAAATPACATTTCTATATAC 980  
DB 656 TTGAAGTACTCGTATACCTGCTTACTCTCAAGTGAAGAAACATGCCATTTCTATATTA 715  
QY 981 TTGACAGACAGATTTGATGCTTGTGCTCAACAGGGTCTGGGAAAGCTGGGCTTTTC 1040  
DB 716 AGGGAAGAAAGACTTATAGCTTGTGTGCCAAGAGATCTGGGAAAGCTGACGATTC 775  
QY 1041 TCTTACCAATTTGGCTCA-----TATGATGATGATG 1073  
DB 776 TTTTACCCATCTAGAGCATATATACAGATGTCAGAGAGAGCTTTGAAGGCTGTGA 835  
QY 1074 GAATTAAGTCAAGCTTTTAAAGATTTGAGAGACAGAGTATTTATTTAGACCAA 1133  
DB 836 AGGAAATGGAAGATATGAGGCGCCGCAACAAATATCCATATCTGTTTATGCCCCAA 895  
QY 1134 CTGAGAAATGCTCAACAGATTTATTTGAGAGCCAGAAATTTCTTTTGGAGCTTGTG 1193  
DB 896 CAAGAGATTTGGCTGTACAGATCTATGAGAGAACCCAGAAATTTCTTCCATCCATCAGAG 955  
QY 1194 TTAAGAGCTGTTTATATATGAGGAGAACCCAGCTGGGAGATTCATTCGACAAATAGTAC 1253  
DB 956 TTGCTCTTGTATGATTTATGAGTGTGCTGATTTGATGACAGATTTGGGACTTGAAC 1015  
QY 1254 AAGGCTGTATATATATGCTACTCTCTGAAAGACTGATGATCATAGGCAAGAAA 1313  
DB 1016 GTGATGCGCACTTGTATGAGCAGCTCCAGAGAGCTTAAGTATATGAAAGAGGAA 1075  
QY 1314 AGATGCTGCAACAGATTAATTAATTTGATGATGATGATGATGATGATGATGATG 1373  
DB 1076 AGATGATTAAGCTTCTGCAAGTACTTATGTTGATAGAGCTGATGATGATGATG 1135  
QY 1374 TGGGTTTGGTCCAGAAATGAAAGATTAATTTCTTGGCCAGAAATGCCATCAAGGAAAC 1433  
DB 1136 TGGGATTTGAACTCAGATACGTCGATAGTTGAAACAGATATCTAGCCAAAGGGCG 1195  
QY 1434 AGGCGCAACCTTATGCTGAGTCAATTTTCCAGAGAAATTCAGAGTTGGCTGCAG 1493  
DB 1196 TTGTGCAACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1255  
QY 1494 AGTTTAAAGTCAAAATTAATCTGTTGTTGCTGTTGAGCAAGGGGTGAGGATGATGAG 1553  
DB 1256 ACTTTTGGATGAA---TATATCTTTTGGCTGTAGGCAAGATGAGCTTACTCTGACA 1312  
QY 1554 ATGTTACAGACAGCGTCTCCAAAGTTGGCAGTTCTCAAAAAAGAAAGAGCTGTGAAA 1613  
DB 1313 ACATTCACAGAGAAAGATTTGGTGGAGAGACTTAAATACGATCAATTTCTACAGACA 1372  
QY 1614 TTCTGGGAAACATAGG---GATGAAAGAACTATGCTTTGTAATAAGAAAGAAAG 1670  
DB 1373 TTTAAGTGCAGAGGAGATGATTAATTAATTTGATGATGATGATGATGATGATGATG 1432  
QY 1671 CAGATTTTACGCACTTTTCTTGTCAAGAAATATCAACTATCAAGTATCAATGATG 1730  
DB 1433 CAGATTCCTGGAGATTTCTTATATACATGAGGATATGCTGTATATATTAATTCATGAG 1492

QY 1731 ATCCGGAACAGAGAGAGCGGAGAGCAAGCTCTTGGAGATTTTCGCTTTGGAAAGTCCAG 1790  
DB 1493 AOCGGTCAACAGAGATGAGAGAGAGGCGCTTACCAAGATTTGCTGATGAGAAAGGCCCA 1552  
QY 1791 TTCTTTGCTTACTTCACTTACCTTCCAGAGGCTCGATTTGAAATGTCACATGTTA 1850  
DB 1553 TTCTAGTGGCTACAGCTGCTGGACACAGAGACTGACATTTCAAAATGATGACATGTTA 1612  
QY 1851 TCAATTTGATCTTCTTCTTCAACATGATGAAATATGATGCAATTTGGGGGTACTGTC 1910  
DB 1613 TCAATTTGATTTTCCCAAGATATTTGAAGAAATATGATGCAATTTGGGGGTACTGTC 1672  
QY 1911 GTTGTGGAATATCGGAGAGCAATTTCTTTTGTATCTTGAATGGAATPACCATTTAG 1970  
DB 1673 GTGTAGGAAACCTGGGCTTGGCCACTCATTTCTTTAA---TGAAGAAATATGAAATTTA 1729  
QY 1971 CACAGCCCTCTGTAAGATATGACAGATGCTCAAGAGATGTTCCCTGATGTTGAGAG 2030  
DB 1730 CAAGGATTTTGTGATCTTCTTGTATGAGGATTAACAGAGAGTCTTCTTGGTTGAAA 1789  
QY 2031 AAATGCTTTTATGATACATATTCCTGCTTCAAGTGTAG 2070  
DB 1790 ATATGCTTATGAACACCACTACAAAGGTGGAGATGCTGG 1829

## RESULT 5

US-10-293-582-4  
; Sequence 4, Application US/10293582  
; Publication No. US20030175253A1  
; GENERAL INFORMATION:

; APPLICANT: Akil, Huda  
; APPLICANT: Burney, William E.  
; APPLICANT: Burke, Sharon  
; APPLICANT: Choudhary, Prabhakara V.  
; APPLICANT: Cox, David R.  
; APPLICANT: Evans, Simon  
; APPLICANT: Jones, Edward G.  
; APPLICANT: Li, Jun  
; APPLICANT: Lopez, Juan F.  
; APPLICANT: The Trustees of The Leland Stanford Junior University  
; TITLE OF INVENTION: Compositions and Methods for Diagnosing and Treating  
; TITLE OF INVENTION: Mental Disorders  
; FILE REFERENCE: 020865-000210US  
; CURRENT APPLICATION NUMBER: US/10/293,582  
; CURRENT FILING DATE: 2002-11-12  
; PRIOR APPLICATION NUMBER: US 60/339,252  
; PRIOR FILING DATE: 2001-11-09  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 4416  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: dead box, y isoform (DBY) RNA helicase  
US-10-293-582-4

Query Match 17.2%; Score 383.2; DB 15; Length 4416;  
Best Local Similarity 58.5%; Pred. No. 1.2e-97;  
Matches 761; Conservative 0; Mismatches 503; Indels 36; Gaps 4;

QY 801 AGACAGCATTAAGTTCGACAAATAGACACTATTCCTGTGGAAAGTGTGGACATGATG 860  
DB 529 ACACGGGGATTTAACTTTGAGAAATATGATATACAGTAGAGGCAACCGGAGTAATCT 588  
QY 861 CACACAGCAATTCCTGACTTTTGAAGAACTAATCTGTGACAGACTGAATPACACA 920  
DB 599 GTCTCCACATATGAAATTTTAGCATATGACATGGGAAATTTATCATGAGGAGACA 648  
QY 921 TTGCTAAAGCTGTTATACCTTAAGCTTAATCTGTGCAAAAAATPACATTTCTATATAC 980  
DB 649 TTGAAGTACTCGTATACCTGCTTACTCTCAAGTGAAGAAACATGCCATTTCTATATTA 708



QY 981 TTGACAGACAGATTTGATGGCTTGTGCTCAACAGAGTCTGGGAGACTGCGGCTTTC 1040  
 Db 709 AGGAAAAAAGACTTATGGCTTGTGCCAAACAGAGTCTGGGAACTGACAGCAATTC 768  
 QY 1041 TCCACCAATTTTGGCTCA-----TATGATGATGATG 1073  
 Db 769 TTTTACCACTAGTACAGATATATATACAGATGTCAGGAGAGAGCTTTGAAGCTGTGA 828  
 QY 1074 GAATTAATGCAAGTCTTTTAAAGAGTTCAGAGAACCAAGTATATATGACCCAA 1133  
 Db 829 AGGAAATGAGAGTATGGCGCGCCAAACATATCCATATCTTGGTTTAAAGCCCAA 888  
 QY 1134 CTGAGAAATGAGTCAACCAAGTATATTTGGAGCCGAAATTTCTTTGGAGCTTGG 1193  
 Db 889 CAGAGAAATTTGGCTGTACAGATCTATAGAGAGCCGAAATTTCTTCCATCCGATCTAGAG 948  
 QY 1194 TAAAGAGTGTGTATATATGAGGGAACCCAGCTGGACATTCATTCGCAAAATGATC 1253  
 Db 949 TTGATCTTGTGTATATGATGAGTGTGATATGATGATGATGATGATGATGATGATG 1008  
 QY 1254 AAGCTGTATATATATATGATGATGATGATGATGATGATGATGATGATGATGATG 1313  
 Db 1009 GTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1068  
 QY 1314 AGATTTGCTCAAAACAGATCAAAATCTTATGATTTGATGATGATGATGATGATGATG 1373  
 Db 1069 AGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1128  
 QY 1374 TGGGTTTGTGCTCAAAAGATGAGAGATTTCTTCCAGAGATGATGATGATGATGATG 1433  
 Db 1129 TGGATTTGAAACCTCAATAGCTGTATGATGAAAGATGATGATGATGATGATGATG 1188  
 QY 1434 AGGCTCAACCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1493  
 Db 1189 TTGCTACACCATATATATGATGATGATGATGATGATGATGATGATGATGATGATG 1248  
 QY 1494 AGTTTAAATCAATATATGATGATGATGATGATGATGATGATGATGATGATGATG 1553  
 Db 1249 ACTTTTGGATGAA---TATATCTTTTGGCTGTAGAGAGATGATGATGATGATGATG 1305  
 QY 1554 ATGTCAGAGACCGTCTTCCAGTGGCCAGTCTCAAAAAGAGAAAGCTGTTGAA 1613  
 Db 1306 ACATCAACAGAAAGTGTGGTGGTGAAGCTTGAATTAACGCTCTTTCTACTGACA 1365  
 QY 1614 TTCTGCAAAACATAGG---GATGAAAGAACTATGATGATGATGATGATGATGATG 1670  
 Db 1366 TTTTATGAGCAACAGGAGGATTCATCTTATGATGATGATGATGATGATGATGATG 1425  
 QY 1671 CAGATTTTACTGCACTTTCTTTGTCAAGAAAAATATCACTCAAGATATCATGATG 1730  
 Db 1426 CAGATTCCTGAGAGATTTCTTATACATGAAGATATGCTGTACATGATGATGATG 1485  
 QY 1731 ATTCGGAACAGAGAGAGCGGAGCAAGCTCTTGAGATTTTGGCTTTGAAAGTGGCCAG 1790  
 Db 1486 ACCGCTCAGAGAGATGAGAGAGAGCGCTTACACAGTTTCCGTGAGGAAAAAGCCAA 1545  
 QY 1791 TTCTTGTGTACTTCACTAGTACAGAGGCTGTGATGATGATGATGATGATGATGATG 1850  
 Db 1546 TTCTAGTGTACAGCTGTGAGAGAGAGATGATGATGATGATGATGATGATGATGATG 1605  
 QY 1851 TCAATTTGATCTTCTTCACTAGTATGATGATGATGATGATGATGATGATGATGATG 1910  
 Db 1606 TCAATTTGATTTGCAAGTATGATGATGATGATGATGATGATGATGATGATGATG 1665  
 QY 1911 GTTGTGGAAATACGAGAGCAATTCCTTTTGTGATCTTGAATCGGATTAACATTAG 1970  
 Db 1666 GTGTAGGAAACCTGGGCTTGGCACTCATCTTTAA---TGAAAAAATATGAAATATTA 1722  
 QY 1971 CACAGCCCTAGTAAAGATATGACAGATGCTCAACAGAGATGCTGCAATGTTGAG 2030  
 Db 1723 CAAAGATTTTGTGATCTTCTTGTAGAGCTTAAACAAAGATGCTTCTTGTGTTGAAA 1782  
 QY 2031 AAATGCTTTAGTACATATCTTCTGCTTCAAGTGTAG 2070

Db 1783 ATATGCTTATGAACACCACTAACAGAGTGTGACATGCTGG 1822

RESULT 6  
 US-09-764-891-1170  
 ; Sequence 1170, Application US/09764891  
 ; Publication No. US20030077808A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rosen et al.  
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 ; FILE REFERENCE: PC006  
 ; CURRENT APPLICATION NUMBER: US/09/764,891  
 ; PRIORITY FILING DATE: 2001-01-17  
 ; Prior application data removed - consult PALM or file wrapper  
 ; NUMBER OF SEQ ID NOS: 10231  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 1170  
 ; LENGTH: 591  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-09-764-891-1170

Query Match 16.0%; Score 356; DB 10; Length 591;  
 Best Local Similarty 99.4%; Pred. No. 1.7e-90;  
 Matches 356; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1867 TTCTACCATGATGATATATGATGATGATGATGATGATGATGATGATGATGATGATG 1926  
 Db 1 TTCTACCATGATGATATATGATGATGATGATGATGATGATGATGATGATGATGATG 60  
 QY 1927 CAGAGCAATTTCTTTTGTGATCTTGAATCGGATTAACATTTAGACAGCTCTTAGTAA 1986  
 Db 61 CAGAGCAATTTCTTTTGTGATCTTGAATCGGATTAACATTTAGACAGCTCTTAGTAA 120  
 QY 1987 AGTATTTGACAGATGCTCAACAGAGATGCTGCTGATGATGATGATGATGATGATG 2046  
 Db 121 AGTATTTGACAGATGCTCAACAGAGATGCTGCTGATGATGATGATGATGATGATG 180  
 QY 2047 ATACATTTCTGCTTCACTGATGATGATGATGATGATGATGATGATGATGATGATG 2106  
 Db 181 ATACATTTCTGCTTCACTGATGATGATGATGATGATGATGATGATGATGATGATG 240  
 QY 2107 AAAAGGCAAGACCTTTGAAACAGAGCTGGTCTTCTTACAGAGCTCCCAATCCAGT 2166  
 Db 241 AAAAGGCAAGACCTTTGAAACAGAGCTGGTCTTCTTCAAGAGCTCCCAATCCAGT 300  
 QY 2167 AGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2224  
 Db 301 AGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 358

RESULT 7  
 US-10-032-585-6564  
 ; Sequence 6564, Application US/10032585  
 ; Publication No. US20030180953A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Terry, Roemer D.  
 ; APPLICANT: Bo, Jiang  
 ; APPLICANT: Charles, Boone  
 ; APPLICANT: Howard, Bussey  
 ; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery  
 ; FILE REFERENCE: 10182-005-999  
 ; CURRENT APPLICATION NUMBER: US/10/032,585  
 ; CURRENT FILING DATE: 2001-12-20  
 ; NUMBER OF SEQ ID NOS: 8000  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 6564  
 ; LENGTH: 2019  
 ; TYPE: DNA  
 ; ORGANISM: Candida albicans  
 US-10-032-585-6564

Query Match 14.7%; Score 327; DB 15; Length 2019;  
 Best Local Similarity 55.8%; Pred. No. 7.6e-82;  
 Matches 699; Conservative 0; Mismatches 530; Indels 24; Gaps 3;

QY CAGGCAATTAAGTGAATTAAGCACTATCTCTGGGAAGTCTGGACATGATGCAC 863  
 DB CAGGCAATTAAGTGAATTAAGCACTATCTCTGGGAAGTCTGGACATGATGCAC 863  
 QY 485 CTGGTATTAAGTGAATTAAGCACTATCTCTGGGAAGTCTGGACATGATGCAC 544  
 DB 485 CTGGTATTAAGTGAATTAAGCACTATCTCTGGGAAGTCTGGACATGATGCAC 544  
 QY 864 CACGAGAAATCTGACTTTTGAAGAGCTATCTCTGGGAAGTCTGGACATGATGCAC 923  
 DB 864 CACGAGAAATCTGACTTTTGAAGAGCTATCTCTGGGAAGTCTGGACATGATGCAC 923  
 QY 545 CAGAACTATTAAGTGAATTAAGCACTATCTCTGGGAAGTCTGGACATGATGCAC 604  
 DB 545 CAGAACTATTAAGTGAATTAAGCACTATCTCTGGGAAGTCTGGACATGATGCAC 604  
 QY 924 CTAAGAGTGAATTAAGTGAATTAAGCACTATCTCTGGGAAGTCTGGACATGATGCAC 983  
 DB 924 CTAAGAGTGAATTAAGTGAATTAAGCACTATCTCTGGGAAGTCTGGACATGATGCAC 983  
 QY 605 AGTTGCTTAAGTGAATTAAGCACTATCTCTGGGAAGTCTGGACATGATGCAC 664  
 DB 605 AGTTGCTTAAGTGAATTAAGCACTATCTCTGGGAAGTCTGGACATGATGCAC 664  
 QY 984 CAGGAGAAATTAAGTGAATTAAGCACTATCTCTGGGAAGTCTGGACATGATGCAC 1043  
 DB 984 CAGGAGAAATTAAGTGAATTAAGCACTATCTCTGGGAAGTCTGGACATGATGCAC 1043  
 QY 665 CTGGTATTAAGTGAATTAAGCACTATCTCTGGGAAGTCTGGACATGATGCAC 724  
 DB 665 CTGGTATTAAGTGAATTAAGCACTATCTCTGGGAAGTCTGGACATGATGCAC 724  
 QY 1044 TACCAATTTTGGCTCATATGATGATGATGATGATGATGATGATGATGATGATGATG 1090  
 DB 1044 TACCAATTTTGGCTCATATGATGATGATGATGATGATGATGATGATGATGATGATG 1090  
 QY 725 TCCAGATTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 784  
 DB 725 TCCAGATTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 784  
 QY 1091 -----TTTAAAGAGTTGCAAGAACCAAGTGTATTAATTAAGCACTATCTCTGGGAAGTCTGG 1145  
 DB 785 CTTTCTCTTCAACAAGTTTATTCACACACTTCTTATTAAGCACTATCTCTGGGAAGTCTGG 844  
 QY 1146 TCAACAAGTTTATTAAGCACTATCTCTGGGAAGTCTGGACATGATGCAC 1205  
 DB 845 TTTCTCAATTTTGAAGAAAGTAAATAATTAAGTATTAAGTATGATGATGATGATGATG 904  
 QY 1206 TTAATTAAGGAGGAAACCACTGAGCACTATTAAGTATTAAGTATGATGATGATGATGATG 1265  
 DB 905 TCGTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 964  
 QY 1266 TATTAAGTGTACTCTGGAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1325  
 DB 965 TGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1024  
 QY 1326 AACAGATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGT 1385  
 DB 1025 CCAACATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGT 1084  
 QY 1386 CAGAAATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGT 1445  
 DB 1085 CACAAATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGT 1144  
 QY 1446 TTAATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGT 1505  
 DB 1145 TGAATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGT 1202  
 QY 1506 CAAATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGT 1565  
 DB 1203 -AATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGT 1261  
 QY 1566 CCGTTTCAAGTGGCCAGTTCTCAAAAGAAAGAAAGTCTGTAATTTTCAAGAAACA 1625  
 DB 1262 AGATCTGTAAGTGAAGTGAAGAAAGAAAGTCTGTAATTTTCAAGTGAAGTGAAGTGAAGT 1321  
 QY 1626 TGGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1685  
 DB 1322 ATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1381  
 QY 1686 CTTTTCTTTGTCAGAAATAATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGT 1745  
 DB 1382 ATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1441  
 QY 1746 AGGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1805  
 DB 1442 AAAGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1501  
 QY 1806 CAGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1865

DB 1502 CCGTTGCTGCAAGAGTGTGATTAATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGT 1561  
 QY 1866 CTTTACATTAAGTGAATTAAGTGAATTAAGTGAATTAAGTGAATTAAGTGAATTAAGTGAATTAAGT 1925  
 DB 1562 CTAGTGAATTAAGTGAATTAAGTGAATTAAGTGAATTAAGTGAATTAAGTGAATTAAGTGAATTAAGT 1621  
 QY 1926 GCGAGCAATTTCTTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1985  
 DB 1622 GTATG 1678  
 QY 1986 AAGTATTAAGTGAATTAAGTGAATTAAGTGAATTAAGTGAATTAAGTGAATTAAGTGAATTAAGT 2038  
 DB 1679 AATTAATTAAGTGAATTAAGTGAATTAAGTGAATTAAGTGAATTAAGTGAATTAAGTGAATTAAGT 1731

## RESULT 8

US-10-437-963-58448  
 ; Sequence 58448, Application US/10437963  
 ; Publication No. US20040123343  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa, Thomas J.  
 ; APPLICANT: Kovalic, David K.  
 ; APPLICANT: Zhou, Yihua  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Wu, Wei  
 ; APPLICANT: Boukharov, Andrey A.  
 ; APPLICANT: Barbazuk, Brad  
 ; APPLICANT: Li, Ping  
 ; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(53221)B  
 ; CURRENT APPLICATION NUMBER: US/10/437,963  
 ; CURRENT FILING DATE: 2003-05-14  
 ; NUMBER OF SEQ ID NOS: 204966  
 ; SEQ ID NO 58448  
 ; LENGTH: 2362  
 ; TYPE: DNA  
 ; ORGANISM: Oryza sativa  
 ; FEATURE:  
 ; NAME/KEY: unsure  
 ; LOCATION: (1)..(2362)  
 ; OTHER INFORMATION: unsure at all n locations  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_601630.1  
 ; US-10-437-963-58448

Query Match 13.6%; Score 301.6; DB 17; Length 2362;  
 Best Local Similarity 54.6%; Pred. No. 1.5e-74;  
 Matches 726; Conservative 0; Mismatches 574; Indels 30; Gaps 5;

QY 801 AGACAGCATTAAGTGAATTAAGTGAATTAAGTGAATTAAGTGAATTAAGTGAATTAAGTGAATTAAGT 860  
 DB 676 AACTGCAATTAAGTGAATTAAGTGAATTAAGTGAATTAAGTGAATTAAGTGAATTAAGTGAATTAAGT 735  
 QY 861 CACGAGCAATTTCTGACTTTTGAAGAGCTATCTCTGAGACATGATTAAGTGAATTAAGTGAATTAAGT 920  
 DB 736 TGCCACCGCAGTCAATTAAGTGAATTAAGTGAATTAAGTGAATTAAGTGAATTAAGTGAATTAAGT 795  
 QY 921 TTGCTTAAGTGAATTAAGTGAATTAAGTGAATTAAGTGAATTAAGTGAATTAAGTGAATTAAGT 980  
 DB 796 TACGAGTGAATTAAGTGAATTAAGTGAATTAAGTGAATTAAGTGAATTAAGTGAATTAAGTGAATTAAGT 855  
 QY 981 TTGAGAGCAATTTGATG 1040  
 DB 856 TTGAGAGCGGAGTCTCATG 915  
 QY 1041 TCTTCAATTTTGGCTCATATG 1100  
 DB 916 GTTCCCTATCATTAAGTGAATTAAGTGAATTAAGTGAATTAAGTGAATTAAGTGAATTAAGTGAATTAAGT 975  
 QY 1101 TGAGG---AACAGAGTATTAAGTGAATTAAGTGAATTAAGTGAATTAAGTGAATTAAGTGAATTAAGT 1157

Db 976 GGAAGTGCATACCTCTTGCATGATTTATCCCCCACTGCGAATATATCATGCCAAATTC 1035  
QY 1158 ATTGGAAAGCCAGAAAATTTCTTTGGAGCTTGTGAAGCTGTGTATATATGAGG 1217  
Db 1036 ATGAAAGAAAGAAATTTGCATACCAAGCTGTGTAGAGGTGTGCTTATGCTG 1095  
QY 1218 GAACCCAGCTGGGACATTCATTCAGCAAAATGACAAAGCTGTATATATATGCTG 1277  
Db 1096 GAGCAACCAATATCATCAAGTTGAGGAATGGAAGGGCGCTGGAATCTGTGGCA 1155  
QY 1278 CTCCTGGAAGCTGATGATATCATATAGCAAAAGAAAGATGATGCTCAACAGATCAAT 1337  
Db 1156 CTCCTGCTGCTGATGATGATCTGCTGAGAGGGCTGAGATATCATCTGCAATGTGAAGT 1215  
QY 1338 ACTTATGTTTGGATGAGCTGATGCTGATGATGATGATGATGATGATGATGATGATG 1397  
Db 1216 ACTTATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1275  
QY 1398 AGTTAATTTCTGCGCCAGATGCCATCAAGAAAGAAAGCCCAACCTTATGCTGATG 1457  
Db 1276 AAATTTGTTGAGCAGATGACATGCCCTCGCGGTGAGGCAAGCAATGTTTATG 1335  
QY 1458 CAACCTTTCCAGAGAAATTCAAAGTTGGCTGAGAGTTTAAAGTCAATATCTG 1517  
Db 1336 CTACTTTTCCAAAGAAATACAGGCGCATGGCTTCAG---TTTCTTGGGACCTACATCT 1392  
QY 1518 TTGTTGCTGTGCAAGTGGGTGAGCATGATGATGATGATGATGATGATGATGATGATG 1577  
Db 1393 TCCTTGTGTGAGAGAGTTGTGTCAAGTCTGATGATGATGATGATGATGATGATGATG 1452  
QY 1578 TTGGCCAGTCTCAAAAGAAAGAAAGCTCGTGAATTTGCG-----G 1620  
Db 1453 TCCTTGAAGCGGATTAAGAAATTAATCTCATGATCTCTTCCATGCCCAAAAGCGAATG 1512  
QY 1621 AAATATAGGGGATGAAAGA---ACTATGCTTTTGTGAACCTAAGAAAGAAAGCAAT 1676  
Db 1513 GTACTTCATGGGAACAGCGCTTACTCTGTCTTGTGAGACAAAGAGGGAGCTGATG 1572  
QY 1677 TTACTGCACTTTCTTTGTCAAGAAATATCAATCAAGATTCATGATGATGATGATG 1736  
Db 1573 CTTTGGAACTGCTTTTACCAAAAGGTTCCCTGACAAAGATTCATGAGGACCGAA 1632  
QY 1737 AACAGAGAGCGGGAGCAAGCTCTTGAAGTTTCCCTTGGAAAGTGCAGATCTTG 1796  
Db 1533 CACACAGGAAGAGAGATGACCGCTTATGATCTTCAAGAGTGAAGCAACCCATCTTG 1692  
QY 1797 TTGCTTACTTCAATGCTGCGAGAGGCTGATATGAAATGTCGAACATGTTATCAAT 1856  
Db 1693 TTGGAGCTGATGATGCTGCGCGGACCTTGCATATCTGATGCTGATGATGATGATG 1752  
QY 1857 TTGATCTTCTTACCATGATGATATGATGATGATGATGATGATGATGATGATGATG 1916  
Db 1753 TTGATCTTCTTACCATGATGATATGATGATGATGATGATGATGATGATGATGATG 1812  
QY 1917 GGAATCTGCGAGAGCAATTTCTTTTATCTTGAATCGATMACCATTTAGCACAGC 1976  
Db 1813 GGAAGTCTGCGCTTGGCAATGATTTTCA---CGAGGCAATCTATCCCTGCGAGGC 1869  
QY 1977 CTCTAGTAAAGATTTGACAGATGCTCAACAGATGCTTCTGATGATGATGATGATG 2036  
Db 1870 CGTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1929  
QY 2037 CCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2096  
Db 1930 CGGCGCGCTTCTTCTTCTGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1989  
QY 2097 TTGATACAG 2106  
Db 1990 TTGGTGGCG 1999

RESULT 9  
US-10-369-493-46077

Sequence 46077, Application US/10369493  
Publication No. US2003023675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xiandeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
PRIOR FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 46077  
LENGTH: 1815  
TYPE: DNA  
ORGANISM: Saccharomyces cerevisiae  
US-10-369-493-46077  
Query Match 13.4%; Score 298.2; DB 16; Length 1815;  
Best Local Similarity 54.2%; Pred. No. 1.2e-73;  
Matches 684; Conservative 0; Mismatches 553; Indels 24; Gaps 3;  
QY 804 CAGGATTAACCTTGCACAAATAGCACTATTTCTGTGAAAGTGTCTGCATATGAC 863  
Db 353 CTGATTAATCACTTCATATCAATGATGATGATGATGATGATGATGATGATGATG 412  
QY 864 CACGCAATCTGACTTTTGAAGAAGCTAATCTCTGACAGACTGATTAACAACATTTG 923  
Db 413 CTGAACCAATCAACAATTTTACTCACTCCATGGAAGATGATGATGATGATGATGATG 472  
QY 924 CTAAAGCTGTATTAATAAGCTTACTCTGTGCAAAATTAAGATTCCTATCATCTTG 983  
Db 473 AATGGCCCGTTTCCACCAAGCAACCTGTGCAAAATTAATCCCTCTTATGTTGCA 532  
QY 984 CAGGAGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1043  
Db 533 ACGGAGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 592  
QY 1044 TACCAATTTGCTCATATGATGATGATGATGATGATGATGATGATGATGATGATG 1085  
Db 593 TCCAGTGTGTCGATATCATTTAAGCTGACATCTCTCAACAGATGTCAGGCT 652  
QY 1086 GTGCTTTTAAAGTTGACAGAACCAAGATGATGATGATGATGATGATGATGATGATG 1145  
Db 653 CCTTTTACCAAGAAAGGCTTACCACTGCTGATATGATGATGATGATGATGATGATG 712  
QY 1146 TCACCAATTTTATTTGAGAGCCAGAAATTTCTTTTGGACTGTGTAAGCTGTG 1205  
Db 713 CCACCAATTTTGCATGAGCCAGAAATTTTACTTATAGATCTGTGTAAGCTGTG 772  
QY 1206 TTATATAGGGGAAACCAAGCTGAGCAATTCATTCAGCAATGATGATGATGATGATG 1265  
Db 773 TCGTACAGTGTGTTCTCCAAATGATGATGATGATGATGATGATGATGATGATGATG 832  
QY 1266 TATTTATGCTCTCTGGAAGACTGATGATGATGATGATGATGATGATGATGATGATG 1325  
Db 833 TTTTATGCTCTCTGGAAGACTGATGATGATGATGATGATGATGATGATGATGATGATG 892  
QY 1326 AACAGTCAATTAATGATTTGATGATGATGATGATGATGATGATGATGATGATGATG 1385  
Db 893 CCAGCTCAAGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 952  
QY 1386 CAGAAATGAAGATTAATTTTTCGCCAGAAATCCATCAAGAGAGAGAGAGAGAGAGAG 1445  
Db 953 CTCAAATTAAGATTAATTTTTCGCCAGAAATCCATCAAGAGAGAGAGAGAGAGAGAG 1012  
QY 1446 TTATGTTCAAGTCAATTTTCCAGAGAAATCAAGGTTGCTGAGAGTTTAAAGT 1505  
Db 1013 TGTATGTTCAAGTCAATTTTCCAGAGAAATCAAGGTTGCTGAGAGTTTAAAGT 1069

QY 1506 CAATTATCTGTTGTTGCTGTGGACAATGGGTGGAGCATTAAGATGTTTACAGCA 1565  
DB 1070 GTGCTACATCTTTTGTCTGTTGGTAGAGTCGTTCTACTCAGAAAACTTACTCAAA 1129  
QY 1566 CCGTTTCCAGTTGGCCAGTCTTCAAAAAGAAAAGCTCGTTGAATTTCCGAAACA 1625  
DB 1130 AAGCTTATACGTTGAAAAATCAAGATTAAGAGTCAAGCTTATTGGATCTATTGTCTGCAT 1189  
QY 1626 TAGGGGATGAAAAGACTATAGCTCTTTGTTGAAACTAAGAAAAAGCAGATTTTACTGCA 1685  
DB 1190 CCACTGCGGTTTGAATCTTATCTTTGTCGAAACTAAGAGATGCGAGATCAATTGACCG 1249  
QY 1686 CTCTTCTTTTCAAGAAAAATATCAACTCAAGTATCAATGTSATCGGAAACAGAGAG 1745  
DB 1250 ATTTCTGATCAATGCAAAATCTTAGAGCTTACCCCACTCATGATGACCGTACCAATCTG 1309  
QY 1746 AGCGGAGCAAGCTCTTGAGATTTTCGTTTGAAAAGTGCACAGTTCTTGTCTACTT 1805  
DB 1310 AGAGAGAACGTGCTTGCGCGCTTCAGATCTGTGCGCGTACTTATTGTTGGCAGAG 1369  
QY 1806 CAGTACGTGCAAGGCTGTGATTAATGAAATGTGCAACATGTTATCAATTTTGAATCTTC 1865  
DB 1370 CTGTGCGAGCTTAAGGCTTAGATTAATTCAAAAGTCAACCGTATCACTAGATTTAC 1429  
QY 1866 CTCTACCATGATGATATATGTCATGAAATTTGGGCTACTGCTGCTGTGGGAATACTG 1925  
DB 1430 CAGTATGTCATGATTAATGTCATAGAAATTTGTGAAGTGTGCTGCGGTAACACCG 1489  
QY 1926 GCGAGCAATTTCTTTTGTGATCTTGATCGATACCAATTTAGACAGCCTCTAATA 1985  
DB 1490 GTCTTGCACTGCTTTTTCACAGTGAAGAAACAGTAC--ATTGTTAAAGGTTTGCAATG 1546  
QY 1986 AAGATTAAGACAGATGCTCAACAGAGATTTCTGATGTTGGAAGAAATTTGCTTATGA 2045  
DB 1547 AATTTTGAAGTGAAGCTTAACCAAGAGTCCCATCATTTTGAAGAGCGTATGATGAGTG 1606  
QY 2046 C 2046  
DB 1607 C 1607

RESULT 10  
US-10-369-493-46167  
Sequence 46167, Application US/10369493  
Publication No. US2003023675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xiandeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 46167  
LENGTH: 1854  
TYPE: DNA  
ORGANISM: *Saccharomyces cerevisiae*  
US-10-369-493-46167

Query Match 13.4%; Score 298; DB 16; Length 1854;  
Best Local Similarity 54.5%; Pred. No. 1,4e-73;  
Matches 676; Conservative 0; Mismatches 540; Indels 24; Gaps 3;

QY 804 CAGGCTAAACTTCGACAAATAGACACTATTTCTTGGAAGTGTCTGGACATGATGAC 863  
DB 389 CTGGTAAAGTTGTAATATAGATAATATTCCTGATGATCATCTGAAAAAGATGTT 448

QY 864 CACGACATTTCTGACTTTTGAAGACATATCTCTGTACAGACTGAATTAACACTTTG 923  
DB 449 CTGAACCTATATCTGATTAATGATCTCCCTCGATGAGCTGTTATAGAAAAATATCA 508  
QY 924 CTAAAGCTGTTTAACTAGCTTACTCTGTGCAAAAATACAGTATCTATCTATCACTTG 983  
DB 509 AACTGTCTAGTTTACCAAACTTACTCAGTACAAAATATTTCTATTCCAAATGTTTCA 568  
QY 984 CAGACAGATTTGATGCTTGTGCTCAAAACAGGCTTGGAAAGACTGCGGCTTTCTCC 1043  
DB 569 AAGGTAGGATTTATATGACATGCGCAAAACGGGCTCAGGTAAAGACAGGTGGTTTGT 628  
QY 1044 TACCAATTTTGGCTCATATGATGATGATG-----AATACTGCA 1085  
DB 629 TTCCGCTTTACGAGAGCTATTAAGGTCTGGCCCTTCCAGTACCGGAAGAACAGAG 688  
QY 1086 GTCTTTTAAAGATTTSCAGGAACACAGATGATTAATTTAGACCAACTCGAAGATTTG 1145  
DB 689 GCTTTTATTAAGAAAGGATATATCATCTGCTTATGACTGCAACCAACTAGAAATTTG 748  
QY 1146 TCAACGATTTATTTGGAAGCCGAAATTTTCTTTGGGACTGTGTAGAGCTGTG 1205  
DB 749 CTACCCAAATTTTGAAGAGCAAGAAATTTGATTAATCTTGGGTTAGGCGATGCG 808  
QY 1206 TTATATATGGGGAAACCCAGCTGGGACATTCAAATTCGACAAATAGTACAAAGCTGTAATA 1265  
DB 809 TTGTTATATGGGCGCCCAATGGTAAACAGATGAAGAGGTTGACCGCGGCTGTGACT 868  
QY 1266 TATATATGCTACTCTCTGGAAGACTGTGATATCATAGCAAAAGAAATTTGTCTCA 1325  
DB 869 TATTTGGGCTTACCCAGGCGGCTGATATCTTATTAAGACCGGAAAGTTTCTTTGG 928  
QY 1326 AACGATCAAAATCTTATGTTTGGATGAAGCTGATGCAATGTTGATGATGTTTGGT 1385  
DB 929 CTATATCAATATCTGATTTGGATGAGCTGATGAATGCTAGATATGGAATTTGGAAC 988  
QY 1386 CAGAAATGAAGATTAATTTCTTGGCCAGGAATGCCATCAAAAGAAACAGCGCCAAACC 1445  
DB 989 CCCAAATTCGCAATATGTTGAAGATGATATGCCCCCTGTGGAATAAGACAGACCT 1048  
QY 1446 TTATGTCAGTGAACCTTTCCAGAGAAATTCAGAAAGTTGGCGCAGACTTTTAAAT 1505  
DB 1049 TGATGTTTCCGCCACATTCAGATGATATTCACGATTTGGGACGCGATTTCTTACA-- 1106  
QY 1506 CAAATTAATCTGTTTGTGCTGTGACAGAGTGGGTGAGCATTAAGATGTTCAAGACA 1565  
DB 1107 -TAATTAATTTTATTCAGTAGAAGAGTGGTCCACTCAGAAAAATATACCGCAA 1165  
QY 1566 CCGTTTCCAGTTGGCCAGTCTCAAAAAGAGAAAAGCTCGTTGAATTTCTGCGAAACA 1625  
DB 1166 GAATATTAATGTCATGATGACATGACAAAAGTGGCTGTGATTTACTGTGCAAG 1225  
QY 1626 TAGGGATGAAAGAACTATGCTTTGTTGAACTAAGAAAAAGCAGATTTTACTGCA 1685  
DB 1226 AGCATTAAGGATGAGATGATTTTGTGTAAGCAAAAGAAATGGCGGATCAACCTCAG 1285  
QY 1686 CTCTTCTTTTCAAGAAAAATATCAACTCAAGTATCCATGTTGATCGGAAACAGAG 1745  
DB 1286 ATTTTGTGATCAATGCAAAATTTCAAGCTACACCAATACATGATGACCGACAGCTG 1345  
QY 1746 AGCGGAGCAAGCTCTGAGATTTTTCGCTTTGGAAGTCCAGTTCTTGTGTAAT 1805  
DB 1346 AAGCTGAACGTGCTTATCTGCTTCAAGCTAACAGTATCTCTGTGCAACAG 1405  
QY 1806 CAGTACGTGCAAGGCTGATTAATGAAATGTGCAACATGTTATCAATTTTGAATCTTC 1865  
DB 1406 CTGTAGCAGGAGAGGTTTGAATATCCGAAATGTCACATGTCATTAATTAAGATTTG 1465  
QY 1866 CTCTACCATGATGATTAATGTTCAATGAAATTTGGGCTGCTGCTGTGGGAATATCTG 1925  
DB 1466 CTTCGACATGATGATGATGCTTCAAGAAATTTGTAAGAGCGCTGCTGTGATACAG 1525

QY 1926 GCAGACAAATTCCTTTTGAATCTTGAATACCATTTAGACACCCCTTAGTAA 1985  
DB 1526 GCCTGCTACTTCAATCTTCAATAGTAA---TAAAGAAATATGGAAGGTTGATGG 1582  
QY 1986 AAGTATGACAGATGCTCAACAGAGATGTTCTGCATGTT 2025  
DB 1583 AATTTTAAATGAAGCAATCAAGAGTTCCCAAGTTT 1622

RESULT 11  
US-10-369-46378  
; Sequence 46378, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-26  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 46378  
; LENGTH: 1908  
; TYPE: DNA  
; ORGANISM: Schizosaccharomyces pombe  
US-10-369-493-46378

Query Match 13.2%; Score 294.6; DB 16; Length 1908;  
Best Local Similarity 56.0%; Pred. No. 1.3e-72;  
Matches 661; Conservative 0; Mismatches 484; Indels 36; Gaps 4;

QY 803 ACAGGCAATTAACCTTGCACAAATAGACATATTTCTTGGAGATGCTGACATGATGCA 862  
DB 436 ACTGGATTAATCTTGAAGATAGATGACATCCCTGTAAAGTTCCGGTGAAGATATC 495  
QY 863 CCACGACCAATTCGACTTTGAAGAAGCTAATCTCTGACAGACTGAATTAACAAT 922  
DB 496 GAGCTGTAAATGATTCACCTTCCCTCTTGAACCTCA---TCTTCTCAAAACAT 552

QY 923 GCTAAGCTGTATTAAGCTTAATCTCTGTGCAAAATACAGTATTTCTATCAACT 982  
DB 553 AACTCTCTGCTACACTCAAGCCCAACCCAGTTCAAAAGAACTCCATTTCTATTTGACAA 612  
QY 983 GCAGAGAGAGATTGAGAGGCTTGTGCTCAACAGAGGCTGGGAAAGCTGGGCTTTCTC 1042  
DB 613 TCCGAGCTGATTTGATGAGCTGTGCGCCAAACGATTCGGTGAAGATCTGCTTTCTTA 672

QY 1043 CTACCAATTTGGCTCATATGATGATGATGAATTAATGCAAGTCTGTTTAAAGTTG 1102  
DB 673 TTCCCTATTTCTTCTGCTTGTGATTAAGGACCTCCCGCTGTCTCTGTTGATCAGAT 732

QY 1103 CAGG-----AACCAAGTATTTATTTAGACCAACT 1135  
DB 733 GCCGGTATGAGTTATGCTCTCGCAAGACCTATCTTACCACTTATTTTGGCTCCACT 792

QY 1136 CGAGATTTGCTCAACAGATTTATTTGAGCCAGAAATTTCTTTTGGACTTGTGA 1195  
DB 793 CGTGAATAGTTTGTCAATTCAGAGGAATCTGTAAATCTGTATCCGTTCTTGGGTG 852

QY 1196 AGAGCTGTTGTTATATATGAGGGAACCCAGCTGGACATTAATTTGCAAAATGTACAA 1255  
DB 853 CGTCCCTGTGCTGTATAGCGGTGCGCATATGCTGTCAAAATCCGCAATTTAGACCA 912

QY 1256 GGCTGTAATATATATATGCTACTCTGGAAGACTGATGATATATAGCAAAAGAAAG 1315  
DB 913 GGTTCGACTTCTTCTGCTACTCTCTGCTGTGTTGTTGTTATATGATCGTGTGT 972

QY 1316 ATGCTCTCAAAAGATCAATTAATCTTGAATGATGATGATGATGATGATGATGATG 1375  
DB 973 ATTCTTTGGCAACATTAATTTTGTGTGTGATGATGATGATGATGATGATGATGATG 1032

QY 1376 GGTTTGGTCCAGAAATGAAGATTAATTTCTTGGCCAGAAATGATGATGATGATGATG 1435  
DB 1033 GGTTCAGCTCTCAATTTGTCACATTTGTGAAGTGTGTCATGACAGAGGTTGAAGAA 1092

QY 1436 CGCCAAACCTTATGTTCAAGTCAACTTTTCCAGAGAAATTCAAAGTTGGCTGACAG 1495  
DB 1093 CGTCAAACTCTCAATTTTGTGACATTTCCCGGTATATCCAACTTTTATAGCTGTGAC 1152

QY 1496 TTTTAAAGTCAATTTATGCTGTTGTTGCTGTTGCAACAAGGAGTGAATGATGAT 1555  
DB 1153 TTTTAAAG--GATATGCTTTCTTGAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1209

QY 1556 GTTCAGACAGCCGTTCTCAAGTTGGCCAGTTCTCAAAAGAAAGAAAGCTGTTGAAT 1615  
DB 1210 ATTACTCAAAAGGCTGTACAGTTGAAAGATTTGAAAGGCTAGTTACTTCTGATAT 1269

QY 1616 CTGCAAAATATGAGGATGAAAG--AACTATGCTTTGTTGAACATAAGAAAGAA 1672  
DB 1270 CTTCACACCTTAACCTCTGAGAGGCTTACCTGTATCTTTGAGAGACTTAAAGCAT 1329

QY 1673 GATTTTACCTGCAACTTTCTTGTCAAGAAATATCAATCAAGATATCATGATGAT 1732  
DB 1330 GACACACTTACCGACTATTTGTTGAATGCAACTCCCGGACAAAGATTCACGGTAC 1389

QY 1733 CGGGAACAGAGAGCGGAGCAAGCTTTGAGATTTTGGCTTTGAAAGTCCCACT 1782  
DB 1390 CGTACCAACGAGAGCGGTGAGCGGCTTTGAAATTTGTCCTGCTGCGGATCTTCAT 1449

QY 1793 CTGTTGCTACTCTGATGCTGACAGAGGCTGATTTGAATGATGATGATGATGATG 1852  
DB 1450 ATGGTGTACTGCGGTGCGCAGTGTGTTAGATTTTCCATGATGATGATGATGATG 1509

QY 1853 AATTTGATCTTCTTCAACATTTGATGATGATGATGATGATGATGATGATGATGATG 1912  
DB 1510 AACTAGATTTTACCTACTGATGATGATGATGATGATGATGATGATGATGATGATG 1569

QY 1913 TGTGGATATCTGAGAGCAATTTCTTTTGTGATCTTGA 1953  
DB 1570 GCTGTAACTGTGCAAGCTGTGCTTTTCAACCTTAA 1610

RESULT 12  
US-10-424-599-14656  
; Sequence 14656, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 14656  
; LENGTH: 3012  
; TYPE: DNA  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_113240C.1  
US-10-424-599-14656

Query Match 13.1%; Score 291.4; DB 13; Length 3012;  
Best Local Similarity 54.3%; Pred. No. 1.5e-71;  
Matches 719; Conservative 0; Mismatches 566; Indels 38; Gaps 5;

QY 769 TGAGATGAGACTCATCTTGTGCAATATATGACAGCAGCATTAACCTTGACAAATACGA 828

Db 749 TAAAGCTGAAGAGGCACTTCACTGAGCAGAGAAATACAGATTAACCTTTAATGCAATGTA 808  
 Qy 829 CACTATTTCTTGTGAAGTGTCTGACATGATGACCAACCAATTTGCACTTTTGAAGA 888  
 Db 809 AGGCACTTGAAGTGAACCAAGTGTGCAATGTCCCACTGTGAAATACGTTTGAGA 868  
 Qy 889 AGCTAATCTGTGAGACATGATTAACATCAATGCTAAAGCTGTGTTACTAGCTTAC 948  
 Db 869 GATTGATTTGGGTAGACACTTAAATCAGAAATATAGAGGTGCAAAATGTGTAAGCAAC 928  
 Qy 949 TCCGTGCAAAAATACAGTATTCCTATCATATCTGACAGACGAGATTTGAGCTTGTG 1008  
 Db 929 ACCGTGTACGGGATGCAATACCGATATCTGATGAGCGGAGATTGATGCTTGTG 988  
 Qy 1009 TCAACAGGGCTGTGGGAAGATGCGGCTTTCTCCATCAATTTGGCTCATATGATGCA 1068  
 Db 989 GCAACTGTGTTCTGGAAAGACAGCTGATTCGTCTCCGATTATAGTGAATCATATGAG 1048  
 Qy 1069 TGATGAATTAATCTGCACTGCTTTTAAAGATTGC-----AGGAACCAAGATGATAT 1122  
 Db 1049 GGGCCAAACCCCTGACAGAGCCACCTGCGGGGTGCGTACAGTGTATCCACTTGCCTTGT 1108  
 Qy 1123 TGTAGCACTCACTGAGAAATTTGCTCAACAGATTTATTTGGAAGCCAGAAATTTTCTTT 1182  
 Db 1109 TCTTTCACCAAGAGAGAGCTATCCATGCAAAATACATGAAGAGCTTAGAAGTTTTCATA 1168  
 Qy 1183 TGGGACTTGTATGAAGAGCTGTTGTTATATATGAGGGAACCCAGCTGGGACATTCATTTG 1242  
 Db 1169 CCAACCTGGGCTTAAGGTGTGTGTGTGATATGCTG-----ACCACTAAATTAACCAAG 1223  
 Qy 1243 ACAAAATGTACAAAGCTGTAAATATATATGCTGCTGAGAGAGCTAGATATCAT 1302  
 Db 1224 GGAATCTTGAAGAGGGGTGAGCACTTCTGTGCACTCTGGAAGAGCTGTAGATTTGCT 1283  
 Qy 1303 AGGCAAAAGAAAAGATTGCTCAACAGATCAAAATCTATGTTTGAAGAGAGCTGATG 1362  
 Db 1284 GGAAGAGCTAGAGTTCACCTGCAATGATTCGATATCTGAGCTTATGATGAGCGAGATG 1343  
 Qy 1363 CATGTTGATATGGGTTTGTGTCAGAAAATGAAGATTAATTTCTTCCAGAGATGCG 1422  
 Db 1344 GATGCTGATATGGGTTTGTGAGCAAAATGAAGAGATTTGTAAGCAATGACATGCG 1403  
 Qy 1423 ATGAAGAGAAAGCCGCAACCTTATGTTCACTGACATTTTCCAGAGAAATTCAAAG 1482  
 Db 1404 TCCACAGAGTCCAGACAGATATGTTGTTCACTGACATTTCCAAAAGAGATTCAGAG 1463  
 Qy 1483 GTTGGCTGAGAGTTTAAAGTCAATATATCTGTTGTGCTGTGGAAGAAGTGGTGG 1542  
 Db 1464 ATGGCTTCTGA---TTTCTTTCAAAATATATTTTCTGCTGTGGAAGAGTGGTTC 1520  
 Qy 1543 AGCATGTAGAGATGTTCAAGACCGTTCCTCAAGTTGGCCAGTTCTCAAAAAGAAA 1602  
 Db 1521 AAGTACTGATTAATTTGTCAAAAGAGTGTAGTATGTTCAAGAGTCTGACAAAGAGATCA 1580  
 Qy 1603 GCTCGTTGAAATTTCTGCG-----AAACATAGGGGATGAAGAAAGAC 1641  
 Db 1581 CCTAATGACCTTCTTCATGACAGAGGGCAATGTGTACAGAGAAAGCAAGCTTTAAC 1640  
 Qy 1642 TATGCTCTTGTGAACTTAAGAAAAGACAGATTTTACTGCAACTTTTCTTGTGAGA 1701  
 Db 1641 TTTAGTTTGTGAGACTAATAGAGAGTGTATTCGCTGGAACACTGTGTGTCTTAA 1700  
 Qy 1702 AAAAATATCACTACAGATCCATGATGATGAGGGAACAGAGAGCGGGAGCAAGCTCT 1761  
 Db 1701 TGGTTTCCGCAACTATCTATGATGAGACAGGTCACAGAGGAAAGAAATTAAGCTTT 1760  
 Qy 1762 TGAAGATTTTCCCTTGGAAAGTCCAGTTCTTGTGCTATCTTCACTAGTGTGCAAGG 1821  
 Db 1761 GAGTCATTTAAAGTGGCAACCCCAATATGTTGTCACCCAGATGTTGCTGACAGTGG 1820  
 Qy 1822 GCTGATATTTGAATGTGCAACATGTATCAATTTGATCTTCTTTCACATTTGATGA 1881

Db 1821 TCTTAATATCTCCCATGTTGCTCATGTGTCAACTTTGACCTTGCCAAATGACATTGATGA 1880  
 Qy 1882 ATATGTTGATCGAATTTGGGCGTACTGTGTGTGGAATATCTGCGAGAGCAATTTGCTT 1941  
 Db 1881 TTATGTACACCGAATTTGAGAGACAGGCGAGCTGGAAGAAAGAGTCTGTGCACTGATT 1940  
 Qy 1942 TTTTATCTTGAATGTGATTAACCATTTTGCACAGCTCTAGTAAAGATTTAGACAGATGC 2001  
 Db 1941 CTTTAA---TGACAAATTTGATGCTGCTAGAGCTTTATCAAGAACTGATGCAAGAGC 1997  
 Qy 2002 TCAACAGATGTTCTGATGATGTTGGAAGAAATTTGCTTGTAGTACATATCTGCTGCTT 2061  
 Db 1998 AAATCAAAAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2057  
 Qy 2062 CAG 2064  
 Db 2058 CAG 2060  
 RESULT 13  
 US-10-424-599-14660  
 ; Sequence 14660, Application US/10424599  
 ; Publication No. US20040031072A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa Thomas J  
 ; APPLICANT: Kovalic David K  
 ; APPLICANT: Zhou Yihua  
 ; APPLICANT: Cao Yongwei  
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with  
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(53223)B  
 ; CURRENT APPLICATION NUMBER: US/10/424,599  
 ; CURRENT FILING DATE: 2003-04-28  
 ; NUMBER OF SEQ ID NOS: 285684  
 ; SEQ ID NO 14660  
 ; LENGTH: 2672  
 ; TYPE: DNA  
 ; ORGANISM: Glycine max  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: PAT\_MR13847\_113244C.1  
 ; US-10-424-599-14660  
 Query Match 12.7%; Score 282.2; DB 13; Length 2672;  
 Best Local Similarity 54.1%; Pred. No. 5,7e-69;  
 Matches 737; Conservative 0; Mismatches 578; Indels 48; Gaps 6;  
 Qy 769 TGAAGATGAGAGCTCCATCTTTGCACTTATGACAGAGGATTAACCTTGACAAATACGA 828  
 Db 537 TAAAGCTGAAGAGCACTTCACTGACAGAGAAATACAGATTAATTTGATGCAATGTA 596  
 Qy 829 CACTATTTCTTGTGAAGTGTCTGACATGATGACCAACCAATTTCTGATTTGAGA 888  
 Db 597 AGACATTCAGTGAAGACAGTGTGACAAATGTGCCCCACCTGTGAATACGTTTGACA 656  
 Qy 889 AGCTAATCTGTGACAGCTGAAATTAACAATTTGCTAAAGCTGTATATCAATGCTTTC 948  
 Db 657 GATTGATTTGGGTGAAGCACTTAATCAAGATTAAGAGGTGCAAAATATGTAAGCAAC 716  
 Qy 949 TCCGTGCAAAAATATACAGTATTCCTATCAATCTTGTGACAGACAGATTTGATGCTTGTG 1008  
 Db 717 ACCGTGTACGGGCAATGCCATACCATATCTGAGTGAACGGAGATTTGATGCTTGTG 776  
 Qy 1009 TCAACAGGGTCTGGAAGAGCTGCGGCTTTTCTCTCAACAAATTTTGGCTCATATGATGCA 1068  
 Db 777 GCAAGCTGTTCTGGAAGACAGCTGCAATTCGCTTCCGATTATCAAGATCATATGAG 836  
 Qy 1069 TGATGAATTAATCTGCACTGCTTTTAAAGATTGC-----AGGAACCAAGATGATAT 1122  
 Db 837 GGGCCAAACCTGACAGAGGCACTGTGGGTGCTAAGATGATCCCTTGCCTTGT 896  
 Qy 1123 TGTAGACCAACTGCGAATTTGGTCAACCAATTTATTTTGAAGCCAGAAAATTTTCTTT 1182  
 Db 897 TCTTCTCCACAGAGGAGTATCAATGCAATATCATGAAGAGGCTTAGAAGATTTTCATA 956

```
QY 1183 TGGGACTTGTGTAAGCTGTGTATATATGAGGAC-----CCAGCTGGGACA 1233
DB 957 CCAACTGGGGTAAAGGTAAGTGTGTATATGTAAGGACCAATTAACAGAGGAGAA 1016
QY 1234 TTCA-----ATTGCAAAATAGTACAAAGGCTGTAATATATATATGCTACTCTCGAAG 1287
DB 1017 TGCAGGGACACTTACACAGTGTCTCAGGGGTGACATCTCTTGTGCAACTCTCGAAG 1076
QY 1288 ACTGATGATATCATAGGCAAGAAAGATTGCTCTCAACAGATCAATATTAGTTT 1347
DB 1077 ACTGATGATATCATAGGCAAGATGATGATGATGATGATGATGATGATGATGAT 1136
QY 1348 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1407
DB 1137 AGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1196
QY 1408 TTGCCCAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1467
DB 1197 GCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1256
QY 1468 AGAGAAATTCAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1527
DB 1257 AAAAGATATACAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1313
QY 1528 TGACAAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1587
DB 1314 TGGAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1373
QY 1588 CTCAAAAGGAAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1626
DB 1374 TGACAAAGGAAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1433
QY 1627 AGGGATGAAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1686
DB 1434 AAAGCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1493
QY 1687 TTTTCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1746
DB 1494 TTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1553
QY 1747 GCGGAGAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1806
DB 1554 AAGGAAATTAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1613
QY 1807 AGTAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1866
DB 1614 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1673
QY 1867 TTCTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1926
DB 1674 AAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1733
QY 1927 CAGAGCAATTTCTTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1986
DB 1734 TCTTGAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1790
QY 1987 AGATATGACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2046
DB 1791 ACTGATGACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1850
QY 2047 ATACATCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2089
DB 1851 TTCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1893

RESULT 14
US-10-369-493-26954
; Sequence 26954, Application US/10369493
; PUBLICATION NO. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 26954
; LENGTH: 1681
; TYPE: DNA
; ORGANISM: Neurospora crassa
US-10-369-493-26954

Query Match 12.4%; Score 276.2; DB 16; Length 1681;
Best Local Similarity 53.2%; Pred. No. 2.1e-67; Indels 29; Gaps 3;
Matches 673; Conservative 0; Mismatches 563;

QY 798 ATCAGACAGCATTAACCTGCAAAATACGACATATCTTGTGGAAGTGTGACATG 857
DB 239 AGCAGACATGATGATCACTTCGAGAGTACGATGACATTCGGTGAAGCTTCGGTGACA 298
QY 858 ATGCAACCCAGCAATCTGACTTTTGAAGAGCTAATCTGTGACAGACTGAAATACA 917
DB 299 AGGTGCGGAGCCGCTTCTTACCTTCAGCAACCCCTCTGTGACACACACCTTCATCCA 358
QY 918 ACATGCTTAAAGTGTATATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAG 977
DB 359 AATTCAGTGTGCGCGCTGCAATGTCCCATCTCCGCTTCAAGAAATCCATCCCATCG 418
QY 978 TACTGAGAGAGATTTGATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1037
DB 419 TATGAGGTGTGCTGATGATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 478
QY 1038 TTCTCTACCAATTTTGGCTATGATGATGATGATGATGATGATGATGATGATGATGAT 1080
DB 479 TCTCTTCCCATCTTCTTCTGATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 538
QY 1081 -----TSCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1133
DB 539 CCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 598
QY 1134 CTGCAATTTGTCAACGATTTATTTGAGCCAGAAATTTCTTTTGGGACTGTG 1193
DB 599 CCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 658
QY 1194 TAAGACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1253
DB 659 TCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 718
QY 1254 AAGCTGTATATATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1313
DB 719 GCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 778
QY 1314 AGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1373
DB 779 GTATCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 838
QY 1374 TGGGTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1433
DB 839 TGGGTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 898
QY 1434 AGGCGCAACCTTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1493
DB 899 ACCGCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 958
QY 1494 AGTTTAAAGTAAATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1553
DB 959 AATTCCTCAAG---GACTACATCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1015
```



QY 1554 ATGTTCAGCAGACCGTTCTCCAGTGGCAGTTCCTCAAGAAAAGAGAAAGCGTTGAAA 1613  
DB 1016 ACATACCCCAAGAGTGGAGTATGAGATATGACAGACCGTCCCTTCTGCACA 1075  
QY 1614 TTTCGCAAAATATGGGAGTAAAGAACTATGCTTTGTTGAAAAGAAAAAGCAG 1673  
DB 1076 TTCTCCACACCCAGCGGAGGTCTCACCTTATTTTCGTTGAGACCAACGCGATGGCCG 1135  
QY 1674 ATTTTACTGCACTTTTCTTTGTCAAGAAAAAATATCACTACAGATATCCATGGTATC 1733  
DB 1136 ATTCGCTGTCCGACTTTCTTATCAACAGAACTTCCCGCTACTCTTCATTCAGGTTACC 1195  
QY 1734 GGGAAACAGAGACCGGAGCAGCACTCTTGAGATTTTCGTTGAAAAGTCCCACTTC 1793  
DB 1196 GCATCAGCGTGAACCGTGAAGCGCGCTTTGAGATGTTCCGCAACGCGTCCGCTATTC 1255  
QY 1794 TTGTTGCTACTTCTAGTAGCTCCAGAGCGCTGATATTAAGAAATGTCACATGTTATCA 1853  
DB 1256 TGGTTGCACTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1315  
QY 1854 ATTTTATCTTCTCTTCTACCATGATGATATGTTTCAATTTGGGCGTACTGCTGTT 1913  
DB 1316 ACTACGATCTCCCTACCGACATGATGATGATGATGATGATGATGATGATGATGATGAT 1375  
QY 1914 GTGGGATACTGAGAGAGCAATTTCTTTTATGATCTTGATGATGATGATGATGATGATGAT 1973  
DB 1376 CCGAAACACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1433  
QY 1974 AGCCTTACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2033  
DB 1434 GCGAGCTTCTTGAATCTCTCAAGAGGCTTACAGAGAGTTCCTGCTTCTTGAAGCA 1493  
QY 2034 TTGCC 2038  
DB 1494 TTGCC 1498

RESULT 15  
US-09-938-842A-521  
; Sequence 521, Application US/09938842A  
; Patent No. US20020160378A1  
; GENERAL INFORMATION:  
; APPLICANT: Harper, Jeff  
; APPLICANT: Kreps, Joel  
; APPLICANT: Wang, Xun  
; APPLICANT: Zhu, Tong  
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
; TITLE OF INVENTION: SAME, AND METHODS OF USE  
; FILE REFERENCE: SRIPI300-3  
; CURRENT APPLICATION NUMBER: US/09/938,842A  
; PRIOR FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: US 60/227,866  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: US 60/264,647  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/300,111  
; PRIOR FILING DATE: 2001-06-22  
; NUMBER OF SEQ ID NOS: 5379  
; SEQ ID NO 521  
; LENGTH: 1902  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-09-938-842A-521

Query Match 12.4%; Score 275.8; DB 9; Length 1902;  
Best Local Similarity 52.8%; Pred No 3e-67;  
Matches 707; Conservative 0; Mismatches 602; Indels 30; Gaps 4;

QY 711 GAGGAGAAAGTAGTATCTCAAGAGCAAAAGTACATACCCCTCTCCACCTG 770  
DB 311 GAAAGGAGAGGTGGAGCGTAGGAGAGTGAAGTAAACCCCTTTGAAATAATATGATTCG 370  
QY 771 AGGATGAGAGCTCCATCTTTGACATTTATCAGACAGGATAACTTGACAAATAACACA 830

DB 371 AACCAAAACAGCTTTTACTGACAGATTAATACGTTATTAATTTGATCCATATGAG 430  
QY 831 CTATCTTTGAGAGTGTCTGACATGATGACACACGCAATTTCTGACTTTGAGAG 890  
DB 431 ATATTCGATGAGACAGTGGGATTAATGTCCTCTCTGTTAATCATTCGAGAGA 490  
QY 891 CTATCTCTGACAGCTGATTAACAATTTGCTTAAAGCTGTTATCTAAGCTTACTC 950  
DB 491 TAGATCTGGGAGGACATGATTAATTAATCCGATGATGCAATATGTTAAGCCACAC 550  
QY 951 CTGTGCAAAAATACAGTATTTCTTATCATCTTGGAGAGCAGATTTGATGCTGTGCTC 1010  
DB 551 CTGTTACAGCTGATCCGATTTCCATTTGCTTGAAGGAGGATTTGATGCTGTGCTC 610  
QY 1011 AAACAGGCTCTGGAGAGCTCGGCTTTCTCTCAACAATTTGCTCATATGATGATG 1070  
DB 611 AGACAGGCTCGGAGAGACAGCTCTTTTGTCTTCAATCATATGAGAAATATGAGAA 670  
QY 1071 ATGGAATATGCGCAG--TGTGTTTAAAGGTTGACAGAGAACAGAGTATATGATG 1127  
DB 671 ATCAGCATGATCAGAGACCCCTGTGTTCAAGAACAGCTTACCTTGTGCACTTATCTCT 730  
QY 1128 CACCAATCTGAGAAATGCTCAACAGATTTATTTGAGAGCCAGAAATTTTCTTTGGGA 1187  
DB 731 CACCAACAGAGAGTGGCAAGTCAATATCATGATGAGGCTTAAAGTTCTCTTACCAA 790  
QY 1188 CTGTGATGAGAGCTGTTGTTATATATGAGGAGAACCCAGCTGGGACATTCATTCGACAA 1247  
DB 791 CTGTGATGAGAGTGTGTTGATATGAGAGAACCTTATTAACAGACCTCGGAGAAC 850  
QY 1248 TAGTACAGGCTGTATATATATATATGATGCTCTCTCTGAGAGATGATGATATGAGCA 1307  
DB 851 TTGAGAGGAGAGTGCATATTTCTGTGAGCAAGCTGTGCTGATTAAGATTTGCTGAGA 910  
QY 1308 AAGAAAGATTTGCTCAACAGATCAATCTTATGTTGATGAGTGAATGCTGATG 1367  
DB 911 GAGCTAAGTCTCAATGACATGATGATGATTTTGAAGCTCTGATGAGGCGGATGAGATG 970  
QY 1368 TGGATATGAGTTTGTGTCAGAAATGAGAAATTTTCTGTCAGAGAAATGCTCA 1427  
DB 971 TTGACATGAGTTTGAACCAATTAAGAAATGTCAGAAATGAGATGCTCTCAC 1030  
QY 1428 AGGAACAGCCCAACCTTATGTTGATGAGCACTTTTCCAGAGAAATTCAGAGTTGG 1487  
DB 1031 GTGAGATTTAGACAGACCTGTTTGTAGTCTCATTTCCAGAGAAATTCAGAGCTCG 1090  
QY 1488 CTGACAGATTTTAAAGTCAATTTATCTGTTGCTGTTGAGCAAGTGGGAGCAAT 1547  
DB 1091 CAGCTGACTTCTAG--CAATTTATATTTTGTGCTGAGTGAAGTGGTTCAAGTA 1147  
QY 1548 GTAGAGATGTTAGACAGACCGTTCTCCAGTTGGCCAGTTCCTCAAAAGAGAAAGCTCG 1607  
DB 1148 CCGATTTATTTGTCAAAGGTTGAGTTGCTCTTCACTGTGCAAAAGAGATCATCTCA 1207  
QY 1608 TTGAATTTGCGAATCATAGGAGTGAAGA-----ACTATG 1646  
DB 1208 TGGACCTGCTTACGCTCAGAGAGAGATGGCATCCAGCAAGCAAGCCCTGACCTTAG 1267  
QY 1647 TCTTTGTTGAATCAAGAAAAAGCAGATTTTACTGCACTTTCTTTGCAAGAAAAA 1706  
DB 1268 TTTTGTGAGACAAAGAGAGAGCTGACTTTTGAAGAAATGTTGTTGATCATCATGAGT 1327  
QY 1707 TATCAACTACAGATATTCATGTTGATGAGGAAACAGAGAGCGGAGCAAGCTTTGAG 1766  
DB 1328 TTCCACCAACCTTCATTCACGTTGAGCAAGACAGAGAGAAAGAGAGTGGATGAGAA 1387  
QY 1767 ATTTGCTTTGGAAGTCCCAAGTCTTGTGCTACTTCAATGAGTGCAGAGGCTG 1826  
DB 1388 CTTTCAAGAGTGGAGAACACGATTTTGTGTTGCACTGATGAGAGAGCTGCTG 1447  
QY 1827 ATTTGAAATGTCACATGTTATCATTTTGAATCTTCTTCAATGATGATGATG 1886

```
Db 1448 ACATGCCCCACGTGGCTCATGTGTCACCTTGCATCTGCCAATGACATTGATGACTATG 1507
QY 1887 TTCTATGGAATTGGGCGTACTGTCTGTGCGGAATATCTGGCAGAGCAATTCCTTTTGG 1946
Db 1508 TCCACCGTATTGGACGACGACGTGCTGGCAATCAGGACTAGCACTGCCTTCTTCA 1567
QY 1947 ATCTTGATCGGATTAACCATTTAGCAACGCGCTTAAGTAAAGTATTGACAGATGCTCAAC 2006
Db 1568 A---CGATGGCAACACTCTCACTGCGCCAGACCGCTCGGTGAGCTGATGCAAGAACTAAC 1624
QY 2007 AGGATGTTCTCTGCATGCTT 2025
Db 1625 AAGAAATCCTCTGAGTGGCT 1643
```

Search completed: July 27, 2004, 13:54:46  
Job time : 2093 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 27, 2004, 05:15:24 ; Search time 166 Seconds  
(without alignments)  
7435.005 Million cell updates/sec

Title: US-09-714-865b-1  
Perfect score: 2224  
Sequence: 1 actgaagcaccacgsggsgg.....aagctctgtgttgatgca 2224  
Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 682709 segs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	493	22.2	504	4	US-09-621-976-2923 Sequence 2923, Ap
2	389.8	17.5	420	4	US-09-833-381-1475 Sequence 1475, Ap
3	386.4	17.4	2451	4	US-09-976-584-786 Sequence 786, App
4	383.2	17.2	2319	3	US-09-058-489-90 Sequence 90, App1
5	383.2	17.2	4416	3	US-09-058-489-17 Sequence 17, App1
6	369.4	16.6	3408	3	US-09-058-489-14 Sequence 14, App1
7	369.4	16.6	5322	3	US-09-183-706-42 Sequence 42, App1
8	358.2	7.1	2365	3	US-09-567-995-42 Sequence 42, App1
9	358.2	7.1	2365	4	US-09-328-352-509 Sequence 509, App
10	329.4	5.8	1191	4	US-09-976-584-213 Sequence 213, App
11	329	5.7	1410	4	US-09-543-681A-1401 Sequence 1401, App
12	127.6	5.6	1830121	4	US-09-557-884-1 Sequence 1, App1
13	125.6	5.6	1830121	4	US-09-643-990A-1 Sequence 1, App1
14	125.6	5.6	1830121	4	US-09-643-990A-1 Sequence 1, App1
15	123.6	5.6	1254	5	PCT-US96-05320A-894 Sequence 894, App
16	123.6	5.6	1830121	4	US-09-557-884-1 Sequence 1, App1
17	123.6	5.6	1830121	4	US-09-643-990A-1 Sequence 1, App1
18	116.8	5.3	3230	4	US-08-961-527-203 Sequence 203, App
19	115.2	5.2	3347	4	US-09-702-705-318 Sequence 318, App
20	115.2	5.2	3347	4	US-09-736-457-318 Sequence 318, App
21	115.2	5.2	3347	4	US-09-614-124B-318 Sequence 318, App
22	115.2	5.2	3347	4	US-09-671-325-318 Sequence 318, App
23	115.2	5.2	3347	4	US-09-589-184-318 Sequence 318, App1
24	115.2	5.2	3825	3	US-09-208-742-3 Sequence 3, App1
25	113.4	5.1	1317	4	US-09-543-681A-195 Sequence 195, App
26	110.4	5.0	45613	4	US-09-596-002-22 Sequence 22, App1
27	110	4.9	1941	4	US-09-328-352-1891 Sequence 1891, App

C	28	109.6	4.9	573	4	US-09-702-705-1384	Sequence 1384, App
C	29	109.6	4.9	573	4	US-09-736-457-1384	Sequence 1384, App
C	30	109.6	4.9	573	4	US-09-614-124B-1384	Sequence 1384, App
C	31	109.6	4.9	573	4	US-09-671-325-1384	Sequence 1384, App
C	32	106	4.8	7363	4	US-08-956-171E-19	Sequence 19, App1
C	33	104.4	4.7	905	4	US-09-976-594-788	Sequence 788, App
C	34	103.2	4.6	1587	4	US-09-134-001C-1425	Sequence 1425, App
C	35	92.2	4.1	1245	3	US-09-318-443-1	Sequence 1, App1
C	36	92	4.1	1365	4	US-09-543-681A-3545	Sequence 3545, App
C	37	91.2	4.1	1368	4	US-09-134-001C-816	Sequence 816, App
C	38	91.2	4.1	66985	4	US-09-596-002-29	Sequence 29, App1
C	39	91	4.1	267	4	US-09-313-284A-3206	Sequence 3206, App
C	40	90.8	4.1	490	4	US-09-222-575-126	Sequence 126, App
C	41	90.8	4.1	490	4	US-09-389-681-126	Sequence 126, App
C	42	90.8	4.1	490	4	US-09-620-405B-126	Sequence 126, App
C	43	90.8	4.1	490	4	US-09-339-318-126	Sequence 126, App
C	44	90.8	4.1	490	4	US-09-433-826B-126	Sequence 126, App
C	45	90.8	4.1	490	4	US-09-604-287A-126	Sequence 126, App

## ALIGNMENTS

RESULT 1									
US-09-621-976-2923									
Sequence 2923, Application US/09621976									
Patent No. 6639063									
GENERAL INFORMATION:									
APPLICANT: Dumas Milne Edwards, J.B.									
APPLICANT: Giordano, J.Y.									
APPLICANT: Jobert, S.									
TITLE OF INVENTION: SEQS and Encoded Human Proteins.									
FILE REFERENCE: GENSET.054PR2									
CURRENT APPLICATION NUMBER: US/09/621.976									
CURRENT FILING DATE: 2000-07-21									
NUMBER OF SEQ ID NOS: 19335									
SOFTWARE: Patent.pm									
SEQ ID NO 2923									
LENGTH: 504									
TYPE: DNA									
ORGANISM: Homo sapiens									
FEATURE:									
NAME/KEY: CDS									
LOCATION: 224..472									
US-09-621-976-2923									
Query Match									
Best Local Similarity 99.8%; Pred. No. 8.9e-134;									
Matches 504; Conservative 0; Mismatches 0; Indels 1; Gaps 1;									
QY	775	TGAGACCTCCATCTTGGACATTATCAGACAGCAATTAACCTTGACAAATAGACACTAT	834						
DB	1	TGAGACCTCCATCTTGGACATTATCAGACAGCAATTAACCTTGACAAATAGACACTAT	60						
QY	835	TCCTTGGAGAGTGTGACATGATGACACACCAAGCAATTCGACTTTGAAAGACTTA	894						
DB	61	TCCTTGGAGAGTGTGACATGATGACACACCAAGCAATTCGACTTTGAAAGACTTA	120						
QY	895	TCCTTGGACACCGAATTAACAATCTGAAGTGTATACCACTTACTCTCT	954						
DB	121	TCCTTGGACACCGAATTAACAATCTGAAGTGTATACCACTTACTCTCT	180						
QY	955	GCAAAATACAGATTCCTATCATCTTACAGACAGATTTGATGCTTGCTCAAC	1014						
DB	181	GCAAAATACAGATTCCTATCATCTTACAGACAGATTTGATGCTTGCTCAAC	240						
QY	1015	AGGCTTGGAGAACTCGCGCTTTCTCTACCAATTTGGCTCATATGATGATGATG	1074						
DB	241	AGGCTTGGAGAACTCGCGCTTTCTCTACCAATTTGGCTCATATGATGATGATG	300						
QY	1075	AATATGCGACAGCTTTTAAAGATTTGAGAAACCAAGATGATTTATTTGACACCAAC	1134						
DB	301	AATATGCGACAGCTTTTAAAGATTTGAGAAACCAAGATGATTTATTTGACACCAAC	360						

QY 1135 TCAGAAATTGTCACACAGATTATTTGGAAGCCAGAAATTTCTTTGGACTGTCT 1194  
DB 361 TCAGAAATTGTCACACAGATTATTTGGAAGCCAGAAATTTCTTTGGACTGTCT 420  
QY 1195 AAGAGCTGTTGTATATATATGAGGAGAACCCAGCTGGGACATTCATTGCAAAATAGTACA 1254  
DB 421 AAGAGCTGTTGTATATATATGAGGAGAACCCAGCTGGGACATTCATTGCAAAATAGTACA 479  
QY 1255 AGGCTGTAATATATATATGCTACT 1279  
DB 480 AGGCTGTAATATATATATGCTACT 504

## RESULT 2

US-09-833-381-1475  
Sequence 1475, Application US/09833381  
Patent No. 6672186  
GENERAL INFORMATION:  
APPLICANT: Robison, Keith E.  
TITLE OF INVENTION: No. 6672186el Nucleic Acid and Protein Homologs  
FILE REFERENCE: 5800-119  
CURRENT FILING DATE: 2001-04-11  
CURRENT APPLICATION NUMBER: US/09/833,381  
PRIOR FILING DATE: 2000-02-29  
NUMBER OF SEQ ID NOS: 2050  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 1475  
LENGTH: 420  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-833-381-1475

Query Match 17.5%; Score 389.8; DB 4; Length 420;  
Best Local Similarity 99.5%; Pred. No. 9,9e-104;  
Matches 391; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 818 GACAATATGACACTATTTCTTGTGGAAGTGTGACATGATGACACACCAATTCG 877  
DB 28 GTCAATATGACACTATTTCTTGTGGAAGTGTGACATGATGACACCAATTCG 87  
QY 878 ACTTTGGAAGAGCTATCTGTGTGACACATGATGATGATGATGATGATGATGAT 937  
DB 88 ACTTTGGAAGAGCTATCTGTGTGACACATGATGATGATGATGATGATGATGAT 147  
QY 938 ACTTAAGCTTACTCTGTGTGCAAAATACAGTATTCCTATCATCTTGCAGAGATTTG 997  
DB 148 ACTTAAGCTTACTCTGTGTGCAAAATACAGTATTCCTATCATCTTGCAGAGATTTG 207  
QY 998 ATGGCTTGTCTCAACAGAGTCTGGAGAGACTGGGGCTTTCTCTCAATTTGGCT 1057  
DB 208 ATGGCTTGTCTCAACAGAGTCTGGAGAGACTGGGGCTTTCTCTCAATTTGGCT 267  
QY 1058 CATATGATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1117  
DB 268 CATATGATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 327  
QY 1118 ATTATGTCAGCAACCTGCAAAATTTGTCACACGATTTATTTGGAAGCCAGAAATTT 1177  
DB 328 ATTATGTCAGCAACCTGCAAAATTTGTCACACGATTTATTTGGAAGCCAGAAATTT 387  
QY 1178 TCTTTGGAGCTTGTGTAAGAGCTGTGTATA 1210  
DB 388 TCTTTGGAGCTTGTGTAAGAGCTGTGTATA 420

## RESULT 3

US-09-976-594-786  
Sequence 786, Application US/0976594  
Patent No. 6673549  
GENERAL INFORMATION:  
APPLICANT: Furness, Michael

APPLICANT: Buchbinder, Jenny  
TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS  
FILE REFERENCE: PA-0041 US  
CURRENT APPLICATION NUMBER: US/09/976,594  
CURRENT FILING DATE: 2001-10-12  
PRIOR APPLICATION NUMBER: 60/240,409  
NUMBER OF SEQ ID NOS: 1143  
SOFTWARE: PERL Program  
SEQ ID NO 786  
LENGTH: 2451  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: Incyte ID No. 6673549 035282CB1  
US-09-976-594-786

Query Match 17.4%; Score 386.4; DB 4; Length 2451;  
Best Local Similarity 58.7%; Pred. No. 2,7e-102;  
Matches 763; Conservative 0; Mismatches 501; Indels 36; Gaps 4;

QY 801 AGACAGCATTAACCTTGACCAATATGACACTATTTCTTGTGGAAGTGTGACATGATG 860  
DB 503 ACACGGGATTAACCTTGACCAATATGATATATACAGTAGAGCAACCGGAGTACT 562  
QY 861 CACCACAGCAATCTGACCTTTTGAAGAGCTATCTCTGACACCTGATTAACACA 920  
DB 563 GTCTCCACATATTTGAATTTTGAAGATTTGACATTTGACATGAGAAATATATATGAGGAGACA 622  
QY 921 TTGCTAAAGCTGTTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 980  
DB 623 TTGAACCTTACCTGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 682  
QY 981 TTGAGAGACAGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1040  
DB 683 AGGAAAAAGAGATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 742  
QY 1041 TCTTACCAATTTGGCTCA-----TATGATCATGATG 1073  
DB 743 TTTTACCAATTTGGCTCA-----TATGATCATGATG 802  
QY 1074 GAATTAATGAGAGTGTGTTTAAAGAGTGTGAGAACCAAGATGATTTATTTGACCAA 1133  
DB 803 AGGTAATGAGAGTGTGTTTAAAGAGTGTGAGAACCAAGATGATTTATTTGACCAA 862  
QY 1134 CTCGAGATTTGTCACACAGATTTATTTGGAAGCCAGAAATTTCTTTGGAGCTGTG 1193  
DB 863 CAGAGATTTGTCACACAGATTTATTTGGAAGCCAGAAATTTCTTTGGAGCTGTG 922  
QY 1194 TAAAGCTGTTGTATATATGAGGAGAACCCAGCTGGGACATTCATTTGACAAATAGTAC 1253  
DB 923 TTGCTCTTGTGTATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 982  
QY 1254 AAGGCTGTAATATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1313  
DB 983 GTGAT 1042  
QY 1314 AGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1373  
DB 1043 AGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1102  
QY 1374 TGGGTTTGTGTCAGAAATGAAAGATTAATTTCTTCCAGAGATTCATTAAGAGAAC 1433  
DB 1103 TGGGTTTGTGTCAGAAATGAAAGATTAATTTCTTCCAGAGATTCATTAAGAGAAC 1162  
QY 1434 AGGCGCAACCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1493  
DB 1163 TTGTCACACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1222  
QY 1494 AGTTTAAAGTCATTAATTTCTTGTGTCAGTGTGAGCAAGTGGGTGAGCATGATGAG 1553  
DB 1223 ACTTTTGTGATGAA--TATATCTTTTGGCTGTGAGGAGAGTATGATGATGATGATGATGATGAT 1279

QY 1554 ATGTTGAGGAGACCGTCTTCCAAAGTTGGCCAGTCTCAAAAAGAGAAAGCTGCTTGA 1613  
DB 1280 ACATCACACAGAAAGATGTTGGGTGAAAGACTTAATAAGCGTCATTTCTACTGAGCA 1339  
QY 1614 TTCTGCGAAACATATAGG---GATGAAGAATATGCTCTTTGTTGAACTAAGAAAAAG 1670  
DB 1340 TTTTAGGTGCAACAGGAGAGTATCTACTACTTATGTTGTTGAGACCAAAAAGGGAG 1399  
QY 1671 CAGATTTTACTGCAACTTTTCTTTGTCAGAAAAATATCACTACAGATATCCATGTG 1730  
DB 1400 CAGATTTCCCTGAGATTTCTTATACATGAAAGATATGCTGTACTATATTCATGAG 1459  
QY 1731 ATGCGGAAACAGAGAGAGCGGAGCAAGCTCTTGAGATTTTGGCTTTGAAAGTGGCCAG 1790  
DB 1460 ACCGGTCACAGAGATCGAGAGAGGCCCTTACACAGTTTGGCTCAGGAAAAAGCCCA 1519  
QY 1791 TTCTTGTGCTACTTCAATAGTCCAGAGAGGCTGATATGAAAATGTGCAACATGTTA 1850  
DB 1520 TTCTAGTGGCTACAGCTGTGGCAGACAGAGACTAAGCAATTCAAATGTGACATGTTA 1579  
QY 1851 TCAATTTGATCTCTCTTCCATGATGATGATATGTTCAATGATTTGGGCGTACTGTC 1910  
DB 1580 TCAATTTGATTTGATCCAGATGATGATGATGATGATGATGATGATGATGATGATG 1639  
QY 1911 GTTGTGGGAATACTGCGAGAGCAATTTCTTTTGTATCTTGAATCGATTAACCATTTAG 1970  
DB 1640 GTGTAGAAACCTGGGCTTGGCACCCTCATTTCTTAA---TGAAAAAATATGATATTA 1696  
QY 1971 CACAGCTCTAGTAAAGTATTTGACAGATCTCAACAGATGTTCTCTGATGTTGAAG 2030  
DB 1697 CAAAGGATTTGTTGATCTTCTTGTAGAAAGCTAAACAGAAAGTCTTCTTGTGTTGAAA 1756  
QY 2031 AAATGGCTTTAGACATACATTCCTGGCTTCAGTGTAG 2070  
DB 1757 ATATGGCTTTAGAACACCACTACAGAGGTGGCAGTCTGG 1796

## RESULT 4

US-09-058-489-90  
; Sequence 90. Application US/09058489  
; Patent No. 6103886  
; GENERAL INFORMATION:  
; APPLICANT: Whitehead Institute for Biomedical Research  
; APPLICANT: Lann, Bruce  
; APPLICANT: Page, David  
; TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of  
; FILE REFERENCE: WHI97-089A  
; CURRENT APPLICATION NUMBER: US/09/058, 489  
; EARLIER FILING DATE: 1996-04-10  
; EARLIER APPLICATION NUMBER: 60/041,877  
; NUMBER OF SEQ ID NOS: 91  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 90  
; LENGTH: 2319  
; TYPE: DNA  
; ORGANISM: Human  
US-09-058-489-90

Query Match 17.2%; Score 383.2; DB 3; Length 2319;

Best Local Similarity 58.5%; Prid No. 2.2e-101;  
Matches 761; Conservative 0; Mismatches 503; Indels 36; Gaps 4;

QY 801 AGACAGCATAACTTGACAAATAGCACTATTTCTTGGAAGTGTGCAATGATG 860  
DB 529 ACAGGGGATTAACCTTGAGAAATATGATATATACAGTAAAGGCAACGGCAGTAACT 588  
QY 861 CACACACGCAATTTCTGATCTTTGAAGAAGTATCTCTGACAGACGCAATTAACACA 920  
DB 589 GTCTCCACATATTTGAGATTTTAGCGATTTGACATGGGAGAAATATATCATGGGGACA 648

QY 921 TTGCTAAAGCTGTTATACATAGCTTACTCTCTGCAAAAATACAGTATTCATATAC 980  
DB 649 TTGAATCTTACTGCTATATCTGCTCTTACTCCAGTGCAGAAAAACATGCCATTCCTATTA 708  
QY 981 TTGACAGACAGATTTGATAGCTTGTGTCAAAAGAGGTCTGGGAAGACTGCGGCTTTTC 1040  
DB 709 AGGGAAGAAAGACTAGTGGCTGTGTGCCCAAGAGATCGGGAAGAAACTGACCATTTTC 768  
QY 1041 TCCATCAATTTGGCTCA-----TATGATGATGATG 1073  
DB 769 TTTTACCATATCTGATGATGATATATACATGATGATGATGATGATGATGATGATGATG 828  
QY 1074 GAATATGCTCCAGTCTTTTAAAGTTGACAGAAACAGAGTATATTTATGACACCA 1133  
DB 829 AGAAAAATGAAAGATGAGGCGCGCCGCAAAACATATCCATATCTTGGTTTACCCCCCA 888  
QY 1134 CTGAGATTTGTGCAACAGATTTATTTGAGAGCAGAAATTTCTTTGGGACTGTG 1193  
DB 889 CAAAGAAATGCTGTACAGATCTATAGAGAGCCAGAAATTTTCTTACCATTTAGAG 948  
QY 1194 TAAGAGCTGTTGATATATATGAGGAGAACCCAGCTGGACATTCATTTGACAAATATGAT 1253  
DB 949 TTGCTCTTGTGATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1008  
QY 1254 AAGCTGTAATATATATATGCTTACTCTGGAAGACTGATGATATCATAGCAAGAA 1313  
DB 1009 GTGATGCTCCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1068  
QY 1314 AGATGCTGCTCAACAGATCAAAATCTAGTTTGTGATGATGATGATGATGATGATGATG 1373  
DB 1069 ACATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1128  
QY 1374 TGGGTTTGTGTCAGAAATGAAAGATTAATTTCTTCCAGAAATGCTCATGAAAGAAC 1433  
DB 1129 TGGGATTTGAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1188  
QY 1434 AGCGCAAAACCTTATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1493  
DB 1189 TTGCTCAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1248  
QY 1494 AGTTTAAAGCTCAAAATATGCTGTTTGTGCTGTTGCAAGAGTGGTGAAGCATGATGAG 1553  
DB 1249 ACTTTTGAAGAA---TATATCTTTTGGCTGATGAGAGTGGCTCTACCTGAGA 1305  
QY 1554 ATGTTGACGACACCTTCTCCAAAGTGGCCAGTCTCAAAAAGAGAAAGCTCTGTTAAA 1613  
DB 1306 ACATCACACAGAAATAGTATTTGGGTGAAAGCTTATGATTAACGGTCAATTTCTACTG 1365  
QY 1614 TTCTGCGAAACATAGG---GATGAAGAATATGCTTTGTTGAACTAAGAAAAAG 1670  
DB 1366 TTTTAGGTGCAACAGGAGAGTATTTCACTTATGTTTGTGAGACCAAAAAGGGAG 1425  
QY 1671 CAGATTTTACTGCAACTTTTCTTTGTCAAGAAAAATATCACTACAGATATCCATGTG 1730  
DB 1426 CAGATTTCCCTGAGAGATTTCTTATACATGAAGATATGCTGTATGATATATGAG 1485  
QY 1731 ATGCGGAAACAGAGAGAGGCGGAGCAAGCTCTGAGATTTTGGCTTTGAAAGTGGCCAG 1790  
DB 1486 ACCGGTCACAGAGATGAGAGAGGCCCTTACCAAGTTTGGCTCAGGAAAAAGCCCA 1545  
QY 1791 TTCTTGTGCTACTTCAATAGTCCAGAGAGGCTGATATGAAAAATGTGCAACATGTTA 1850  
DB 1546 TTCTAGTGGCTACAGCTGTGGCAGACAGAGACTGATTTCAATATGTGACATGTTA 1605  
QY 1851 TCAATTTGATCTTCTTCTTCACTTATGATGATGATGATGATGATGATGATGATGATG 1910  
DB 1606 TCAATTTGATTTTGCAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1665  
QY 1911 GTTGTGGGAATACTGCGAGAGCAATTTCTTTTGTGATCTTGAATGATGATGATGATG 1970  
DB 1666 GTGTAGAAACCTGGGCTGTGCACTATCTTTAA---TGAAAAAATATGATATTA 1722  
QY 1971 CACAGCTCTAGTAAAGATTTGACAGATGCTCAACAGAGATGCTCTGATGTTGAGAG 2030

Db 1723 CAAGAGTTTGGTGGATCTTCTGTAGAGCTAAACAGAGAGCTTCTGTGGTGA 1782  
Qy 2031 AAATGCTTTAGTACATATATCTTCGGCTTCAGTGTAG 2070  
Db 1783 ATATGGCTTATGAACACCACTACAGAGGTGGAGTGTGG 1822

RESULT 5  
US-09-058-489-17

/ Sequence 17, Application US/09058489  
/ Patent No. 610386  
/ GENERAL INFORMATION:  
/ APPLICANT: Whitehead Institute for Biomedical Research  
/ APPLICANT: Lahn, Bruce  
/ APPLICANT: Page, David  
/ TITLE OF INVENTION: Genes in the No. 610386-Recombining Region of  
/ TITLE OF INVENTION: the Y Chromosome  
/ FILE REFERENCE: WH197-08PA  
/ CURRENT APPLICATION NUMBER: US/09/058,489  
/ EARLIER FILING DATE: 1998-04-10  
/ EARLIER APPLICATION NUMBER: 60/041,877  
/ EARLIER FILING DATE: 1997-04-11  
/ NUMBER OF SEQ ID NOS: 91  
/ SOFTWARE: FastSeq for Windows Version 3.0  
/ SEQ ID NO 17  
/ LENGTH: 4416  
/ TYPE: DNA  
/ ORGANISM: Human  
US-09-058-489-17

Query Match 17.2%; Score 383.2; DB 3; Length 4416;  
Best Local Similarity 58.5%; Pred. No. 3.2e-101;  
Matches 761; Conservative 0; Mismatches 503; Indels 36; Gaps 4;

Qy 801 AGACAGGCAATTAACCTTCAGCAATATAGACATATCTTGTGGAAGTGTGACATGATG 860  
Db 529 ACAGGGGATTAACCTTGAAGAAATATGATGATATACATAGACCAACGGCAATTA 588  
Qy 861 CACACACAGCAATTCCTGACTTTGAAAGACTTAATCTCTGACACACTGAATTAACA 920  
Db 589 GTCCCTCACATATGAGAAATTTAGGATTTGACATGAGAGAAATTAATCATGGGACA 648  
Qy 921 TTGCTAAAGTGGTATTAAGCTTACTCTCTGTGAAATAATACAGTATTCCTATCATAC 980  
Db 649 TTGAATCTTACTGCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 708  
Qy 981 TTGACAGACGAGATTTGATGCTGTGCTCAACAGAGTCTGGAGAGCTGGGCTTTTC 1040  
Db 709 AGGAAAAAGAGCTAGTGGCTGTGGCCCAACAGATCTGGAAAACTGCAGATTTTC 768  
Qy 1041 TCCCTACCAATTTGGCTCA-----TATGATGATGATG 1073  
Db 769 TTTTACCCATCTGAGTCAAGATATATACAGATGTTCCAGAGAGACTTTGAAGGCTGGA 828  
Qy 1074 GAATTAATCCAGTCTGTTTAAAGATTGACAGAAACAGAGTGTATTTATGTAGACCAA 1133  
Db 829 AGGAAATGGAAGTATGGGCGCCGCAACATATCCATATCTTGGTTTAAAGCCCAA 888  
Qy 1134 CTGAGAAATTTGCTCAACAGATTTATTTGGAAGCCAGAAATTTCTTTTGGAGCTTGT 1193  
Db 889 CAGAGAAATTTGCTGTACAGATCTATGAGAGAGCCAGAAATTTCTTCCATCGATTTAG 948  
Qy 1194 TAAGAGCTGTGTATATATGAGGAGAACCCAGCTGGAGCATTCATTTGACAAATATAGAC 1253  
Db 949 TTGCTCTTGTGTATGTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1008  
Qy 1254 AAGGCTGTATATATTTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1313  
Db 1009 GTGAGTGCACACTGTATAGACCACTCCAGAGCTGTGTGTGTGTGTGTGTGTGTGTGT 1068  
Qy 1314 AGATTGGTCTCAAAAGATCAATACTAGTTTGGATGAGAGCTGATGCGATGTTTGATA 1373

Db 1069 AGATTGATTAGACTTCCAGATPACTTAGTGTGATGATGAGGATGATGATGATA 1128  
Qy 1374 TGGTTTTTGTCCAGAAATGAAAGATTAATTTCTTCCAGAGATGCCATCAAGAGAAC 1433  
Db 1129 TGGATTTGAACCCACATAGCTGTATATGTGAACAAGATPACTATGCCCACAAAGGCG 1188  
Qy 1434 AGCCCAAAACCTTATGTCAATGCACTTTTCCAGAGAGAAATTCAGAGGTTGGCTGAG 1493  
Db 1189 TTGCTACACCATATATTTAGTGTCTCTTTCTTAAGAGAAATACAGATCTGCTGTG 1248  
Qy 1494 AGTTTAAAGTCAATTAATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1553  
Db 1249 ACTTTTGGATGAA---TATATCTTTTGGCTGTAGGACAGATGAGGCTTACTCTGAGA 1305  
Qy 1554 ATGTTGACAGACCGTCTCCAGATGAGCCAGTGTCCAAAGAGAGAAAGCTGTGAGAA 1613  
Db 1306 ACATCAACAGAAAGTATGTTGGTGTGAACACTAGATTAACGTCTATTTCTACTGACA 1365  
Qy 1614 TTCTGCAAAACATAGG--GATGAAAGAACTATGTCTTTGTGAAACTAGAAAAAG 1670  
Db 1366 TTTAGGTGCAACAGGAGTGTATCTTACTTGTAGTGTGTGTGTGTGTGTGTGTGTGTGT 1425  
Qy 1671 CAGATTTTACTGCACTTTCTTTGCAAGAAATATCACTAGATGATGATGATGATGATGAT 1730  
Db 1426 CAGATTTCTGAGAGATTTCTTATACCATGATGATGATGATGATGATGATGATGATGAT 1485  
Qy 1731 ATCGGAAACAGAGAGACGGGAGCAAGCTCTTGAATTTTCGCTTTGAAAGTGGCCAG 1790  
Db 1486 ACCGTACAGAGAGATGAGAGAGGCTTCAACAGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1545  
Qy 1791 TTCTTTGTCTACTTCAATGCTGCGAGAGGCTGTGATTTGAAATGTGCAACATGTTA 1850  
Db 1546 TTCTAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1605  
Qy 1851 TCATTTTGTATCTTCTTCAACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1910  
Db 1606 TCATTTTGTATTTGCCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1665  
Qy 1911 GTTGTGGAATPACTGCGAGAGCAATTTCTTTTGAATCTTGAATGATGATGATGATGAT 1970  
Db 1666 GTGAGAGAAACCTGGGCTTGCACCTCAATCTTTAA---TGAAAAAATAGAAATATTA 1722  
Qy 1971 CACAGCTCTAGTAAAGATTAAGTACAGATGCTCAACAGAGTGTCTGATGATGATGATGAT 2030  
Db 1723 CAAGAGTTTGTGATCTTCTGTGAAAGCTTAAACAAGAGTGCCTTCTGTGTGAAA 1782  
Qy 2031 AAATGCTTTAGTACATATATCTTCGGCTTCAGTGTAG 2070  
Db 1783 ATATGGCTTATGAACACCACTACAGAGGTGGAGTGTGG 1822

## RESULT 6

US-09-058-489-14  
/ Sequence 14, Application US/09058489  
/ Patent No. 610386  
/ GENERAL INFORMATION:  
/ APPLICANT: Whitehead Institute for Biomedical Research  
/ APPLICANT: Lahn, Bruce  
/ APPLICANT: Page, David  
/ TITLE OF INVENTION: Genes in the No. 610386-Recombining Region of  
/ TITLE OF INVENTION: the Y Chromosome  
/ FILE REFERENCE: WH197-08PA  
/ CURRENT APPLICATION NUMBER: US/09/058,489  
/ EARLIER FILING DATE: 1998-04-10  
/ EARLIER APPLICATION NUMBER: 60/041,877  
/ EARLIER FILING DATE: 1997-04-11  
/ NUMBER OF SEQ ID NOS: 91  
/ SOFTWARE: FastSeq for Windows Version 3.0  
/ SEQ ID NO 14  
/ LENGTH: 3408  
/ TYPE: DNA  
/ ORGANISM: Human  
US-09-058-489-14



Query Match 16.6%; Score 369.4; DB 3; Length 3408;  
 Best Local Similarity 57.6%; Pred. No. 2.9e-97;  
 Matches 758; Conservative 0; Mismatches 521; Indels 36; Gaps 4;

786 TCTTTGACATTTACAGACAGGATTAACCTTCAGCAAAATCGACACTATTTCTTGAGAG 845  
 1305 TCTTTTCTGGAGGCAACACTGGGATTTATTTTGAAGAAATGCAATGACATTTCCAGTTGAGG 1364  
 846 TGTCTGGACATGATGACACCAACCAATTCGACTTTTGAAGAGCTAATCTCTGACAG 905  
 1365 CAACAGGCAACACTGTCCTCCACATATTTGAAGTTTCACTGATGTTGAGATGGAGAAA 1424  
 906 CACTGAATTAACAATTGCTTAAAGCTGTTTACTAAGCTTACTCTGTCGCAAAAATACA 965  
 1425 TTATCATGGGAAACATTGAGCTTACTGCTTACTGCTCCAACTCCAGTGCAGAAAGCATG 1484  
 966 GTATTCTCATCATCTTGGACAGAGAGATTTGATGCTTGCTCAACAGGGCTTGAGG 1025  
 1485 CTATTTCTTATTTAAGAAAGAAAGAGACTTGATGCTTGCTGCCAAACAGGGCTTGAG 1544  
 1026 AGACTGGGCTTTTCTCTTCAACATTTTGGCTCATATGATGATG----- 1074  
 1545 AAACCTGACGATTTCTGTTCCCATCTTGAGTCAGATTTATTCAGATGCTCCAGGGAGG 1604  
 1075 -----AATACTGCCAGTGGCTTTTAAAGAGTTGCGAGAACCAAGATGTA 1118  
 1605 CTTTGAGGGCCATGAGAAATGAAAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1664  
 1119 TTATTGAGCAACCACTCGAAGATTTGCTCAACAGATTTTATTTGGAAGCAGAAATTTT 1178  
 1665 TGTATTTAGCCACCAACAGAGAGTTGGCAGTACGATCTCAAGAAAGCCAGAAATTTT 1724  
 1179 CTTTGGGACTTGTCTAAGAGCTTGTATATTTGGGGGAAACCCAGCTGGGACATTTCA 1238  
 1725 CATCCCATCTAGAGTTGCTGCTGCGTGTATATGTTGTCGAGATTTGGTTCAGACA 1784  
 1239 TTGCAAAATAGTCAAGGCTGTATATATATATGCTGCTGCTGCAAGAGCTAGTATA 1298  
 1785 TTCCAGACTTGAACGTCGATGCCATTTGTTAGTACGATCTCCAGAGAGCTGATGATA 1844  
 1299 TCATAGGCAAGAAAGATTTGCTCAACAGATCAAAATTTATGTTTGGATGAGCTG 1358  
 1845 TGATGGAAGAGGAAAGATTTGATTTAGCTTTTGCAAATCTTGCTGTATGATGACCTG 1904  
 1359 ATCCAGTTGGAATATGAGGTTTGTGTCAGAAATGAAAGATTAATTTTGGCCAGGAA 1418  
 1905 ATCGAGTTGGAATATGAGGTTTGTGTCAGAAATGAAAGATTAATTTTGGCCAGGAA 1964  
 1419 TGCCATCAAGAAACAGGCGCAACCCCTTATGTTCACTGCACTTTTCCAGAGAAATTC 1478  
 1965 TGCTCCCAAGAGGAGTCCGACACATGATGTTAGTCTAATTTTCTTCAAGGAAATAC 2024  
 1479 AAAGTTGGCTGCAAGATTTTAAAGTCAATTAATGTTTGGCTGCTGGAACAATGG 1538  
 2025 AGATGCTGGCTGCTGATTTCTTAAGTGA--TAAATCTTCTTGGCTGGAAGAGATG 2081  
 1539 GTGAGCATGTAAGATGTTCAAGACAGCCGTTCTCAAGTTGAGCAGTTCTCAAAAAG 1598  
 2082 GCTTACTCTGAAACATCAACAGAAAGATTTGGGTGGAAGATCAAGACAAAGGT 2141  
 1599 AAAAGCTGTTGAATTTCTGGAACATAG--GATGAAAGAACTATGCTCTTTGTTG 1655  
 2142 CATTTCTGCTTGAACCTCTTAATATGCAACAGGCAAGGATTTCACTAGCTTATGTTTGG 2201  
 1656 AAACTAAGAAAAACAGATTTTACTGCACTTTTCTTCTCAAGAAAAAATATCACTA 1715  
 2202 AGACCAAAAAAGGTCAGATTTCTGAGAGATTTCTTATACCAAGAAAGATACGAGATA 2261  
 1716 CAAGTATCATGATGATGAGGAAACAGAGAGCGGAGCAAGCTCTTGAGATTTTGCT 1775  
 2262 CCAGCATCATGAGACCGTTCTCAGAGGATAGAGAGAGGAGGAGGAGGAGGAGGAGGAG 2321

1776 TTGGAAGTCCCACTTCTTGTCTACTAGTAGTCCAGAGGGGTGATTTGAAA 1835  
 2322 CAGGAAAAAGCCATTTTATGAGCTACAGAGTAGACGAAAGAGTGCATTTCAA 2381  
 1836 ATGTCAACATGTTATCAATTTTGTATCTTCTCTACCATTTGATGATATGTTCAATGAA 1895  
 2382 ATGTAAACATGTTATCAATTTTGTATCTTCTCTACCATTTGATGATATGATATGATGCTA 2441  
 1896 TTGGGCTGATGCTGTTGTGAGAAATCTGCAAGAGCAATTTCTTTTTCATCTGAT 1955  
 2442 TTGTGCTGATGAGGAGCTGTAGAAACCTTGCTGCGCACTCATCTTTAA--CGAGA 2498  
 1956 CGGATACATTTAGCAGAGCTCTAGTAAAGATTTAGCAGATGCTCAGAGATGTTTC 2015  
 2499 GGAACATTAATTTCTAAGATTTTGTGATCTTCTTGTGAAGCTAAACAGAGATGC 2558  
 2016 CTGATGTTGGAAGAAATTTGCTTTAGTACATACATTTCTGGCTTCAGTGTAG 2070  
 2559 CGTCTGTTAGAAACATGCTTATGAAACCACTACAGGGTAGCAGTGTG 2613

RESULT 7  
 US-09-058-489-13  
 ; Sequence 13, Application US/09058489  
 ; Patent No. 6103886  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Whitehead Institute for Biomedical Research  
 ; APPLICANT: Lahn, Bruce  
 ; APPLICANT: Page, David  
 ; TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of  
 ; FILE OF INVENTION: the Y Chromosome  
 ; FILE REFERENCE: WH197-08PA  
 ; CURRENT APPLICATION NUMBER: US/09/058,489  
 ; EARLIER FILING DATE: 1998-04-10  
 ; EARLIER APPLICATION NUMBER: 60/041,877  
 ; EARLIER FILING DATE: 1997-04-11  
 ; NUMBER OF SEQ ID NOS: 91  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 13  
 ; LENGTH: 5322  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 US-09-058-489-13

Query Match 16.6%; Score 369.4; DB 3; Length 5322;  
 Best Local Similarity 57.6%; Pred. No. 3.8e-97;  
 Matches 758; Conservative 0; Mismatches 521; Indels 36; Gaps 4;

786 TCTTTGACATTTACAGACAGGATTAACCTTCAGCAAAATCGACACTATTTCTTGAGAG 845  
 1305 TCTTTTCTGGAGGCAACACTGGGATTTATTTTGAAGAAATGCAATGACATTTCCAGTTGAGG 1364  
 846 TGTCTGGACATGATGACACCAACCAATTCGACTTTTGAAGAGCTAATCTCTGACAG 905  
 1365 CAACAGGCAACACTGTCCTCCACATATTTGAAGTTTCACTGATGTTGAGATGGAGAAA 1424  
 906 CACTGAATTAACAATTGCTTAAAGCTGTTTACTAAGCTTACTCTGTCGCAAAAATACA 965  
 1425 TTATCATGGGAAACATTGAGCTTACTGCTTACTGCTCCAACTCCAGTGCAGAAAGCATG 1484  
 966 GTATTCTCATCATCTTGGACAGAGAGATTTGATGCTTGCTCAACAGGGCTTGAGG 1025  
 1485 CTATTTCTTATTTAAGAAAGAAAGAGACTTGATGCTTGCTGCCAAACAGGGCTTGAG 1544  
 1026 AGACTGGGCTTTTCTCTTCAACATTTTGGCTCATATGATGATG----- 1074  
 1545 AAACCTGACGATTTCTGTTCCCATCTTGAGTCAGATTTATTCAGATGCTCCAGGGAGG 1604  
 1075 -----AATACTGCCAGTGGCTTTTAAAGAGTTGCGAGAACCAAGATGTA 1118  
 1605 CTTTGAGGGCCATGAGAAATGAAAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1664  
 1119 TTATTGAGCAACCACTCGAAGATTTGCTCAACAGATTTTATTTGGAAGCAGAAATTTT 1178

Db 1665 TGGTATTAGCAACGACGAGAGTGGCAGATCTCAGAAAGGACGAGAAATTTT 1724  
Qy 1179 CTTTGGGACTTGTGTAAGCTGTGTATATATGAGGGAACCCAGCTGGGACATTCA 1238  
Db 1725 CATACCATCTAGAGTTCCTTCTGCTGTTATATGAGTGGCCGATATGGTCAGACA 1784  
Qy 1239 TTGACAAATATGATCAAGGCTGTATATATATATGAGTGGTACTCTGGAAGACTGATGATA 1298  
Db 1785 TTCCAGACTTGGAACTGAGATGCGATTTGTAGAGCCACTCCGAGAGCTCTAGTGATA 1844  
Qy 1299 TCATAGGCAAGAAAGATTTGCTCAACAGATCAATCTATGTTTGGATGAAGCTG 1358  
Db 1845 TGAATGAAAGAGAAATGATGATTAAGCTTTGCAATATCTGTTGATGAAGCTG 1904  
Qy 1359 ATGCATGTTGATATGAGTGTGTGTCAGAAATGAAAGATTAATTTCTGCGCAGAA 1418  
Db 1905 ATCGATGTTGATATGAGGTTTGAAGCTTCAAGTTCGTAATAGTGAACAGATATCA 1964  
Qy 1419 TGGCATTAAGGAACAGCGCCAAACCTTATGTTCACTGCACTTTTCCAGAGAAATTC 1478  
Db 1965 TGCTCCAAAGGAGTCCGCACTATGATGTTATGTTGCTACTTTTCTTAAGAAATTC 2024  
Qy 1479 AAGGTTGCTGACAGATTTTAAAGCAATATCTGTTGTTGCTGTTGCAAGTGG 1538  
Db 2025 AAGTGTGCTGCTGATTTCTTAGATGA---TATATCTTGTGCTGATGAGAGGTTG 2081  
Qy 1539 GTGAGCATGTAGAGATGTTCAAGACCGCTTCCAGTGGCCAGTTCTCAAAAAG 1598  
Db 2082 GCTCTACCTCTGAAATCATACACAGAAATGATTTGGTGGAAAGATGACAAACGCT 2141  
Qy 1599 AAAAGTCTGTTAAATCTGCAAAACATAG---GGATGAAGAACTATAGTCTTGTG 1655  
Db 2142 CATTTGCTTGAACCTCTAAATGCAACAGGCAAGATTCACGACCTTAGTGTG 2201  
Qy 1656 AAACCTAAGAAAAAGCAGATTTTACTGCACTTTTCTTGTCAAGAAAAATATCACTA 1715  
Db 2202 AGACCAAAAAGGCTGAGATTTCTGAGAGATTTCTTATACATGAAGATACGATGA 2261  
Qy 1716 CAAGTATCCATGATGATCGGGAACAGAGAGCGGAGACAGTCTTGAAGATTTTGGT 1775  
Db 2262 CCGACATCATGAGACCGTCTCAGAGGATGAGAAAGAGCGCTTCCAGATTCGCT 2321  
Qy 1776 TTGGAAGTCCAGATTTCTGTTGCTACTTCACTAGTCCAGAGGCTGATTTGAAA 1835  
Db 2322 CAGGAAAAAGCCCAATTTTAGTGGCTACAGAGTACAGAAAGAGACCTGGAATTTCA 2381  
Qy 1836 ATGTGCAACATGTTATCAATTTGATCTTCTTCACTTGAATATGTTCAATGAA 1895  
Db 2382 ATGTGAAACATGTTATCAATTTGATCTTCCAAATGATATGAAAGATATGACATGTA 2441  
Qy 1896 TTGGGCGTACTGCTGTTGAGGAATAGTGGCAGAGCAATTCCTTTTGTGATTTGAT 1955  
Db 2442 TTGGTGTACGGGACGTGAGAGAACTTGGCTGCAACCTCATTTCTTTAA---CGAGA 2498  
Qy 1956 CGGATTAACATTTAGACAGCTCTTAGTAAAGATTTGACAGATGCTCAACAGATGTT 2015  
Db 2499 GGAACATTAATATATTAAGATTTGTTGATCTTCTGTTGAACCTAAACAGAGGCT 2558  
Qy 2016 CTGCATGTTGGAAGAAATGCTTATGATACATATCTGAGCTTCAAGTGTAG 2070  
Db 2559 CGCTTGTGTTAGAAACATGGCTTATGAACACCACTACAGGCTGACAGTCTGG 2613

RESULT 8  
US-09-183-706-42  
; Sequence 42, Application US/09183706  
; Patent No. 6245525  
; GENERAL INFORMATION:  
; APPLICANT: Martelange, Valrie  
; APPLICANT: De Smet, Charles  
; APPLICANT: Boon-Falleur, Thierry  
; TITLE OF INVENTION: TUMOR ASSOCIATED NUCLEIC ACIDS AND USES THEREFOR

FILE REFERENCE: L0461/7054  
; CURRENT APPLICATION NUMBER: US/09/183,706  
; CURRENT FILING DATE: 1998-10-30  
; EARLIER APPLICATION NUMBER: 09/122,989  
; EARLIER FILING DATE: 1999-07-27  
; NUMBER OF SEQ ID NOS: 43  
; SEQ ID NO 42  
; LENGTH: 2365  
; TYPE: DNA  
; ORGANISM: H. sapiens  
; NAME/KEY: CDS  
; LOCATION: (208)...(2151)  
US-09-183-706-42  
Query Match 7.1%; Score 158.2; DB 3; Length 2365;  
Best Local Similarity 50.1%; Pred. No. 9.2e-36;  
Matches 518; Conservative 0; Mismatches 488; Indels 27; Gaps 4;  
Qy 915 ACAACATTGCTAAGCTGTTATCTAAGCTTACTCTGCAAAAATACAGTATTCCTA 974  
Db 971 AAAACATTAAAAAGGCAAGTTCATAAAGCCAAACCTATTCACTACAGGATGCGCA 1030  
Qy 975 TCATACTGCAAGACAGATTTGATGCTGTGCTCAACAGGCTCGGGAAGCTGCGG 1034  
Db 1031 TTGTTGCAAGAAATGATCTTATAGAGATGCGCACTGGAACGAAAGCATTTGT 1090  
Qy 1035 CTTTCTCTCAACAAATTTGCTCATATGATGATGATGAAATTAATCCAGTCTGTTTA 1094  
Db 1091 GTTATTTATCTCTGATTTATATCATCTGTCCTGCTCAAC-----CGAGCTTAAAG 1141  
Qy 1095 AAGAGTTCAGAGAACAGAGTATTTATGAGACCAATCGAGATTTGTCACACAGA 1154  
Db 1142 GTCAAGGAATAGACCGGCAATGTTAGTTCTAATCCACCTCGGAAATGACATTCAG 1201  
Qy 1155 TTTATTTGAAGCCAGAAAAATTTCTTTGGGACTTGTGTAAGCTGTGTTATATATG 1214  
Db 1202 TAGAAGAGAAATGTGCAAAATTTCAAT---AAGGCTTCGAGAGTGTGTATATG 1258  
Qy 1215 GGGGAACCCAGCTGGACATTCATTTGACAAATAGTACAGGCTGATATATATATG 1274  
Db 1259 GTGCTGGAATATAGATGAAACAAATGAAAGCTTAAAAAGGTATGATATATATATG 1318  
Qy 1275 CTACTCTGGAAGCTGATGATATCATAGCAAGAAAGATTTGGCTCAACAGATCA 1334  
Db 1319 CAATCTCCGGAATATATATGATCTGCAATGATATCTTCTCAATCTGAAGAAATATA 1378  
Qy 1335 AATACCTAGTTTGTGATGAAAGCTGATGCAATGTTGATATGAGTGTGTTGTCAGAAATGA 1394  
Db 1379 CTTACTTGTGTTTATGATGAAGCAGACAAAGATGTTGACATGGGATTTGAACCCAGATTA 1438  
Qy 1395 AGAAGTAAATTTCTGCGCCAGGAATGCCATCAAGAAACAGCGCCAAACCTTATATGTA 1454  
Db 1439 TGAAGATTTTGT-----TAGATGTGCGCCAGATGAGCAAGATTAATATATG 1486  
Qy 1455 GTGCACTTTTCCAGAGAAATTCAAAGTGTGCTGAGAGTTT---AAGTCAAAAT 1511  
Db 1487 GTGCTATAGCGCTCATATGATTCATGCGCTTCACAAATCTTATTTGAAGAACAATGA 1546  
Qy 1512 ATCTGTTTGTGCTGTTGCAAGTGGTGGAGACATGTAAGATGTTTACGACAGACCTTC 1571  
Db 1547 TTGCTATGTTGATCACTTGGATCTAGTGTCTGATGTTCAAGGAAAGAAATATATATG 1606  
Qy 1572 TCCAGTTGCGCAGTTCTCAAAAAGAGAAAGCTGCTGAAATTTGCAAAACATAGAGG 1631  
Db 1607 TAAACACCGAGAGAGAAATGAGATCATGTAACATTTCTACAGAGATATGCTATCA 1666  
Qy 1632 ATGAAGAACTATGCTGTTTGTGTAACATTAAGAAAAAGCAGATTTTACTGCAACTTTC 1691  
Db 1667 CAGACAAAGCATTTGCTGTTTCTGAAAGAGCTGTTCCGAGTCACTTATCAAGTAC 1726  
Qy 1692 TTTGTCAAGAAAAATATCACTACAGATATCATGATGAGGGAACAGAGAGCGGG 1751

Db 1727 TAATCTTGGAATATATCTAGTAGCTCTGCATGAGATAGAGAACAGAGATCGG 1786  
QY 1752 AGCAAGCTCTGGAGATTTTGGCTTGGAAAGTGCCAGTTCTGTGCTACTTCACTAG 1811  
Db 1787 AGAAGCATATAGAACTTTAAACAGGCAAGTGAGAAATCAATTGCACTGATCTAG 1846  
QY 1812 CTGCGAAGAGGCTGGAATATGAAAATGTCACATGTTATCAATTTGTAATCTCTCTTA 1871  
Db 1847 CCTGTAGAGACTGATGTCATGACGTACACATGTCATTAATTTGACTTTCACGGA 1906  
QY 1872 CCATTGATGATATGTTTCATGCAATTTGGCGCTACTGCTGCTTGGGAATACTGCGAG 1931  
Db 1907 ATATTGAAGATAGCTACACCAATAGCGGACCGGAAAGCAGGAGAGACTGATTT 1966  
QY 1932 CAATTTCCTTTT 1944  
Db 1967 CCATTACACTT 1979

## RESULT 9

US-09-567-995-42  
; Sequence 42; Application US/09567995  
; Patent No. 6303756  
; GENERAL INFORMATION:  
; APPLICANT: Martelange, Val,rie  
; APPLICANT: De Smet, Charles  
; APPLICANT: Boon-Falieur, Thierry  
; TITLE OF INVENTION: TUMOR ASSOCIATED NUCLEIC ACIDS AND USES THEREFOR  
; FILE REFERENCE: L0461/7054  
; CURRENT APPLICATION NUMBER: US/09/567,995  
; CURRENT FILING DATE: 2000-05-10  
; PRIOR APPLICATION NUMBER: 09/183,706  
; PRIOR FILING DATE: 1998-10-30  
; NUMBER OF SEQ ID NOS: 43  
; SEQ ID NO 42  
; LENGTH: 2365  
; TYPE: DNA  
; ORGANISM: H. sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (208)...(2151)  
US-09-567-995-42

## Query Match

Best Local Similarity 7.1%; Score 158.2; DB 4; Length 2365;

Matches 518; Conservative 0; Mismatches 488; Indels 27; Gaps 4;

QY 915 ACAACATTGCTMAAGCTGTTTACTAAGCTTCTCTGCAAAAATACAGTATTCCTA 974  
Db 971 AAAACATTTAAAGGCGGTTTTCAAAAGCCACACCTATTTCAGTCACAGCATGCCA 1030  
QY 975 TCATCTTGAGAGAGAGATTTGATGCTGCTCAACAGGCTGAGAGACTGCG 1034  
Db 1031 TTGCTGTCAGAGAGATAGATCTTATAGAGTAGCCAGATCGAAGCAAGAAACACTT 1090  
QY 1035 CTTTCTCCCTACCAATTTTGGCTCATATGATGATGATGAGAAATACCTGCCAGTCGTTTA 1094  
Db 1091 GTTATTTTAATGCTCGATTAATTCATCTGCTCTTCAAC-----CCAGCTTAAAG 1141  
QY 1095 AAGAGTTCAGAGAAACAGAGATTTATTTAGACCACTGCAAGATTTGTCACAG 1154  
Db 1142 GTCAAGAGATAGACCCGCGATGTAGTTCTAATCTCCCATCTCGGAATTAGCACTTCA 1201  
QY 1155 TTTATTTGAAAGCCAGAAATTTCTTTGAGACTGTGTAAGAGCTGTTGTAATATG 1214  
Db 1202 TAGAAGAGATATGCAAAATATTCATAT--AAAGGCTTCGAGAGTTTGTATATG 1258  
QY 1215 GGGAGACCCAGCTGAGACATTCATTCGACAAATAGTACAAAGCTGTAATATTTATG 1274  
Db 1259 GTGCTGAAATAGAGATGACAAATAGAGAGCTTAAAGAGGTGATGATATCATATTTG 1318  
QY 1275 CTACTCTGGAAGACTGATGATATCATAGCAAAAGAAAGATGTGCTCTCAACAGATCA 1334

## RESULT 10

US-09-328-352-509  
; Sequence 509; Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 509  
; LENGTH: 1191  
; TYPE: DNA  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-509

Query Match 5.8%; Score 129.4; DB 4; Length 1191;  
Best Local Similarity 47.6%; Pred. No. 1.6e-27;  
Matches 521; Conservative 0; Mismatches 556; Indels 18; Gaps 4;

QY 880 TTTTGAAGAGCTATCTCTGTGACACAGTGAATTAACAACATTTGTAAGCTGTATAC 939  
Db 51 TTTTGAAGCTTGAATTTTACATCCGCAACTTAAAGGCGATGTGATGCTTAAAGGTTAC 110

QY 940 TAACTTACTCTGTGCAAAAATACAGTATTCCTATCAATCTTGACAGACGAGATTGAT 999  
Db 111 TCAATGACCCCAATTCAGCAAAAAGTTTAAATATACATTAAGCCGGGACATGATCAAT 170  
QY 1000 GGCTGTGCTCAAAAGGGGTCTGGGAAGACTGGGCTTTTCTTCTCAATTTTGGCTCA 1059  
Db 171 TGGGCGGACAGACGAGAAAGGTAAAGACTGCTGCTTCTTGATTAAGTAAATTAATA 230  
QY 1060 TATGATGATGATGATTAATCACTCCAGTCTTTTAAAGAGTTGACGAAACGAGGTAT 1119  
Db 231 TTGCTCAATATCCGCTCAGAGACAGCTTCCGTGGT-----GAGCTCTGCTTT 284  
QY 1120 TATTGAGCAACAATCTGAGAAATTTGTCACACGATTTTATTGGAAGCCAGAAATTTTC 1179  
Db 285 AATCTTGAACCTTCTGATGATTTGATGATGATGATGATGATGATGATGATGATGAT 344  
QY 1180 TTTTGGACTTGTGTAAGAGCTGTTTATATATGAGGGAACCCAGCTGGACATTCAT 1239  
Db 345 AAAATTTCTAATTTACACTAGTACCCATTAAGTGGTGTGATTTGATGATGATGAT 404  
QY 1240 TCGACAAATATGTAAGAGCTGT---AATATATATGCTACTCTCTGGAAGACTGATGA 1296  
Db 405 GAAACCACTGATGCTAATTTTGTGATATATATGTTGCAACACGAGGCGTTTATTTGA 464  
QY 1297 TATCATAGGCAAGAAAAGATTGCTCAAAACAGATCAATCTTATGTTTGGATGAAGC 1356  
Db 465 TTTTGTGAAACAAAAGAAAGTTTGGCTGCATCAATTTGAATTTTATGATTAATGATGAC 524  
QY 1357 TGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1416  
Db 525 TGACCGTTTATGGAATATGAGGCTTTTATCTCTGCTGTAACGATATGCTGATTTCA-- 582  
QY 1417 AATGCCATCAAGGAACAGGCGCAACCCCTATGTTCAAGTCACTTTCCAGAGGAAT 1476  
Db 583 ----CCAGTAAAGAAACAGCTCAAACTTAATGTTCTCTGCAACATTTAGTTAGATGT 638  
QY 1477 TCAAGGTTGCTGCAAGATTTTAAAGTCAATATCTGTTGTTGCTGTTGCAACT 1536  
Db 639 CTGAACTTGGCAAGACAGTGT---TTGAAACAGTAACTGTTGAATTAAGACTGA 695  
QY 1537 GGGTGAAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1596  
Db 696 ACAAAAGCAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 755  
QY 1597 AGAAAAGCTGTTGAATTTCTGGAACATAGGGGATGAAAGATAGTGTCTTTGTTGA 1656  
Db 756 TCGCTTTTACAGATATTTTACGTGAAGCCAAATGATTAAGATATGATCTTTGCCAA 815  
QY 1657 AACTAAGAAAAAGCAGATTTTACTGCACTTTCTTTCTCAAGAAAAATATCACTAC 1716  
Db 816 TCGCCGTGATCAGGTACGTGCTTTATGATGATGATGATGATGATGATGATGATGAT 875  
QY 1717 AAGTATCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1776  
Db 876 GATGCTATCTGTTGAATTTCTCAAGTAAAGCTTTAAATGTTAGACAGTTTAAAGA 935  
QY 1777 TGAAGATGCTGCTTGTGTTACTTCACTAGTCTGCAAGGGCTGGATTTGAAAA 1836  
Db 936 AGGCAAAACATTAATCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 995  
QY 1837 TGGGCAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1896  
Db 996 TGTATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1055  
QY 1897 TGGGCTACTGCTGTTGGAATATCTGGCAGAGCAATTTCTTTTGTATCTTGAATC 1956  
Db 1056 TGGTTCGACAGCTGCTGACAGGGCACAAGGTGTAAGTATTAAGTTCTTATCGAAGATGA 1115  
QY 1957 GGTATACATTTAGC 1971  
Db 1116 TGCTTTCTATTAC 1130

RESULT 11  
US-09-976-594-213  
; Sequence 213, Application US/09976594  
; Patent No. 6673549  
; GENERAL INFORMATION:  
; APPLICANT: Furness, Michael  
; APPLICANT: Buchbinder, Jenny  
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS  
; FILE REFERENCE: PA-0041 US  
; CURRENT APPLICATION NUMBER: US/09/976,594  
; PRIOR FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: 60/240,409  
; NUMBER OF SEQ ID NOS: 1143  
; SOFTWARE: PERL Program  
; SEQ ID NO 213  
; LENGTH: 3760  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURES:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. 6673549 2173757CB1  
US-09-976-594-213  
Query Match 5.8%; Score 129; DB 4; Length 3760;  
Best Local Similarity 48.4%; Pred. No. 3,9e-27;  
Matches 539; Conservative 0; Mismatches 535; Indels 39; Gaps 5;  
QY 844 AGTGTCTGACATGATGACACACACAGCAATTTGACTTTTGAAGAACTATCTGTCTCA 903  
Db 1236 AGTTAAAGAAAGAGTTGCTCCCAACCAATTAATCTGAGTCAAGTGAATTTCCAT 1295  
QY 904 GACACTGAATTAACAACATTTGCTAAAGCTGTTATCTACTACTCTCTGCAAAAATA 963  
Db 1286 GAAATCTTAATTTCCCTCAAGAAAGATGCTATGAAAGCCAGCCCATCCAAACCA 1355  
QY 964 CAGTATCTTATCATCTTCAAGACAGATTTGATGCTGTTGCTCAACAGGCTGTG 1023  
Db 1356 AGTATTCCTGCTAATATGCTGACGACGATTTGATTTGATTTGCCAAACAGGAGTGG 1415  
QY 1024 GAAATCTGCGCTTTTCTCTACCAATTTTGGCTCATATGATCATGATGATGATGATGAT 1083  
Db 1416 AAGAACCAATGCTTTTCTGTTGCCATGTTAAGACATATGATCA----- 1463  
QY 1084 CAGTGTGTTTAAAGATTTGAGGAACCAAGATGATTTATTTGATGACCAACTGAGAAAT 1143  
Db 1464 GAGGTCAATTAGAGGAAGAGAGGGCCAAATAGCTGTCATCATGATCAACTGAGAACT 1523  
QY 1144 GGTCAACCAATTTATTTGGAAGCCAGAAATTTTCTTTTGGACCTTGTATAGACTGT 1203  
Db 1524 GGTCTTACAGATTACTTAAAGGTGTAAAGATTTTCCAGACTTTGGACTTGAAGTGT 1583  
QY 1204 TGTATATATGAGGGAACCCAGCTGGGACATTAATTCACAAATAGTACAAAGCTGTAA 1263  
Db 1584 CTGTGTTTACGAGGAACAGGAATCATGAGCAGATTTGCTGAACTGAAAGAGTCTGA 1643  
QY 1264 TATATATGCTACTCTCTGGAAGACTGATGATATCATAGGCAAAAGAAAGTTGCTCT 1323  
Db 1644 AATATATGTTTGGACACCTGCTGGAATGATGATGATGATGATGATGATGATGATGAT 1703  
QY 1324 CAACAGATC-----AATATCTAGTTTGGATGGAAGCTGATGCTGTTGATAT 1374  
Db 1704 CACAAATCTTGAAGATGACATATGTTTATGATGAGACAGACAGAAATTTTGAAT 1763  
QY 1375 GGGTTTGTGTCAGAAATGAAGATTAATTTCTTCCAGGAATGCCATCAAGAGACA 1434  
Db 1764 GGGTTTGAACCCAGGTCAATGCGCATGCGATTAATGTTGCT-----CCTGA 1811  
QY 1435 GCGCAACCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1494  
Db 1812 TCGACAGACGTTATGTTTCACTACTTCCCAAGCTATGAGAGCTTTGCTGCGAG 1871  
QY 1495 GTTTTAAAGTCAATATCTGTTTGTGCTGTGGAACAAGTGGTGAAGATGATGAGA 1554

Db 1872 GATCTCA---GTAACCTATTGAAAGTGAAGTGGAGGAGAGTGTGTTGCTCAGA 1928  
QY 1555 TGTTCAGCAGACCGCTTCTCCAGAGTTGGCCAGTTCTCAAAAAGAGAAAAGCTCGTTGAAT 1614  
Db 1929 TGTGACCAACAAGATGATTGATTGAAGAAAAGAAATCTTGAAGTACTTGAGCT 1988  
QY 1615 TCTGCCAACAATGAGGAGATGAAGAAGT---ATGCTGTTGTTGAAGTGAAGAAAAG 1671  
Db 1989 TCTAGGCCATTATCAAGATGCAAGATCTGTCATTATTTGTGGATAGCAGGAACATGC 2048  
QY 1672 AGATTACTGCACTTTCTTTGTCTCAAGAAAAATATCAACTACAGTATCCATGTGA 1731  
Db 2049 TGATGCTCTTAAAGATTAAATGAGATCTTATCTTGATGCTCTTCAATGAGG 2108  
QY 1732 TCGGGAACAGAGAGAGCGGAGCAAGCTTTGAGATTTTCGCTTGGAAAGTCCAGT 1791  
Db 2109 CATGATCAATATGACAGAGATGACATCAATGATGATTAAGAAATGGACCTGCAACT 2168  
QY 1792 TCTGTGCTACTTCAGAGCTGCGAGAGGCTGGATATGAAAATGTGCAATGTTAT 1851  
Db 2169 TCTGTGCTACTTCAGAGCTGCGAGAGGCTGGATATGAAAATGTGCAATGTTAT 2228  
QY 1852 CAATTTGANTCTCTTCAACATGATGAAATGTTCAATGCAATTTGGGCGTACTGCTG 1911  
Db 2229 AAATTAAGCTGCCCCCAACCATTTATGAGATTTATGACAGAGAGGCGCGACTGGAAG 2288  
QY 1912 TTGTGGAAATCTGCGAGAGCAATTCCTTTT 1944  
Db 2289 AGCAGAAACAGGTTATGCTTATCTTTAT 2321

RESULT 12  
US-09-543-1401  
; Sequence 1401, Application US/09543681A  
; Patent No. 6605709  
; GENERAL INFORMATION:  
; APPLICANT: GARY BEBTON  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709, 1002-001  
; CURRENT APPLICATION NUMBER: US/09/543,681A  
; CURRENT FILING DATE: 2000-04-05  
; PRIOR APPLICATION NUMBER: US 60/128,706  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 8344  
; SEQ ID NO 1401  
; LENGTH: 1410  
; TYPE: DNA  
; ORGANISM: Proteus mirabilis  
US-09-543-681A-1401

Query Match 5.7%; Score 127.6; DB 4; Length 1410;  
Best Local Similarity 47.1%; Pred. No. 5.8e-27;  
Matches 514; Conservative 0; Mismatches 554; Indels 24; Gaps 3;

Db 880 TTTTGAAGACCTAATCTCTGTGACAGCTGAATTAACAACCTTCTAAAGCTGTATAC 939  
QY 21 TTTTACATCGCTTGAATTAAGTGAAGCGCTTCTCCGCGCTATGATGAACAAGGGTATTA 80  
Db 940 TAAGTCTCTCTGCAAAATACAGTATCTCTATCTACTTGCAGAGCAGATTTGAT 999  
QY 81 AACCCCAACCCCTATTCACAAACAGGCGATTTGAGCCATTTTGGCGGTAAGACGTAT 140  
Db 141 AGCAGTGCACAAACGGGTACGGGAAAAACGGCGCAATTTACATTACCGATACGAGAAA 200  
QY 1060 TATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1119  
Db 201 ACTC-----GCTACATCAAGCAGAGAAAAAGGCGTAAAGCAGTAAAGCGCT 251  
QY 1120 TATGTAGCAACCACTCGAAGATTGGTCAACCAAGATTATTTGGAAGCCAGAAAAATTTTC 1179

Db 252 TATTTTAAACCCCTACACCGTGAATTTAGCTGACCAATTTGCCATTAACATTAAGCTTATAG 311  
QY 1180 TTTTGGAGCTGTGTGAAGAGCTGTGTGTATATATGTGGGAAACCCAGCTGGACATTTCAAT 1239  
Db 312 TCGTATTTTAACTTATTCCTGCTATGCTGCTTTTGGTGGCGTTAGTATCAATTCCTCAAT 371  
QY 1240 TCGACAAATAGTACAGAGCTGTATATATATATATATATATATATATATATATATATATAT 1299  
Db 372 GATGAACATCAAGAGTGGCGTGTATGTTTATGATGCAACCGGCGCCTTACTGATCT 431  
QY 1300 CATAGCAAAAGAAAGATTTGCTCAAAACAGATCAATATCTTATTTGATGAAGCTGA 1359  
Db 432 TGAATCACTAAATGCTGTCACTATCTGCGTTGAAATTTGTTAGATGAAGCTGA 491  
QY 1360 TCGCATGTTGATATGAGGTTTGTGCTCAAGAAATGAAGATTAATTTCTTGGCCAGAT 1419  
Db 492 TCGTATGTTGATATGAGGATTTATTCATGATA-----TTGCTGAGTATCA 539  
QY 1420 GCCATCAAGAGAGAGCGCCAAACCTTATGTTCAAGTGGCCAGTTCTCAAAAGAGA 1479  
Db 540 TAAGTTACGGAAGAAACAGCAAAATTTACTCTTTTCAAGGACTTTTCAAAAGAGATAC 599  
QY 1480 AAGTTGCTGCAAGCTTTTAAAGTCAATATCTGTTTGTCTGTTTGAAGAGTGG 1539  
Db 600 AAGGCT--TGCTTACTCACTACCAACATCTATTAAGTATGCTGTGCCCCCAAAA 656  
QY 1540 TGAGCATGTAGAGATGTTCAAGACCGTCTCAAGTGGCCAGTTCTCAAAAGAGA 1599  
Db 657 CTCAGGCGTGAATCTGTTGATCAATATGTCATTTAGTGAATTAAGCGTAAGACGA 716  
QY 1600 AAGCTGCTGAAATTTCTGCAAACTAGGGGATGAAGAACTATGCTTTGTTGAAC 1659  
Db 717 GCTGTATACACTTAATTTGCTTGAATAATTTGCTCAAGTTCTTATTTTACTCGAAC 776  
QY 1660 TAAGAAAAAGAGATTTTACTGCAACTTTTCTTGTCAAGAAAAATATCAATACACAG 1719  
Db 777 TAAACATGCTGCAATTAATTAAGTGAACATTTTAAATTTAAATGCGATTAATCAGCGCG 836  
QY 1720 TATCCATGCTGATCGGGAACAGAGAGCGGAGCAAGCTCTTGGAGATTTTGGCTTTGG 1779  
Db 837 GATCATGTTAATTAAGCCAGAGAGCTGAACAGCTGCTTAACTGATTTTAAAGACGG 896  
QY 1780 AAGTCCCAATTTCTTGTCTTACTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1839  
Db 897 TAAATTAAGGCTTATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 956  
QY 1840 GCAACATGTTATCAATTTGATCTTCTCTTCACTTCACTTCACTTCACTTCACTTCACTT 1899  
Db 957 TCTTATATGCTGCACTTTGAATTAACGAGGTAGCAGAAAGATTAAGTCACTGATTTGG 1016  
QY 1900 GGTACTGCTGCTGTTGGAATTAAGTGAAGCAATTTCTTTTGAATTTGAATTTGAAT 1959  
Db 1017 TCGAAACAGGCGCTGCGCGCAACAGGTAAACGATATCTTTAGTTGTGTTGATGAACA 1076  
QY 1960 TAACCATTTAGC 1971  
Db 1077 TGGCTTATTAAGC 1088

RESULT 13  
US-09-557-884-1/c  
; Sequence 1, Application US/09557884  
; Patent No. 6506581  
; GENERAL INFORMATION:  
; APPLICANT: Fleischmann et al.  
; TITLE OF INVENTION: The Nucleotide sequence of the Haemophilus influenzae Rd Genome, fragments thereof, and Uses Thereof  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue

Query Match	5.6%;	Score 125.6;	DB 4;	Length 1830121;
Best Local Similarity	47.3%;	Pred. No. 1.3e-24;		
Matches 506;	Conservative	0;	Mismatches 534;	Indels 30;
				Gaps 3

Qy	881	TTTGAAGAAGCTATCTCTGTGACACCTGATATAACAATTGGCTAAAGCTGTAACT	940
Db	444736	TTTGAACAAATTGGATTTTCTCTGTAGCTTTAAAGCACTTGAGAAAAAGTTATCC	444677
Qy	941	AAAGCTTACTCTGTGCAAAAATACGATTTCTTATCATCTTGACAGACGAGATTGATG	1000
Db	444676	CGCCCAACGCAATTCMAATGGAAACCATTCCTGCGCAATGGAAAGAGTAGTGTTA	4446171
Qy	1001	GCTTGTGCTCAACAACGAGCTGTGGAAAGACTGCGGCTTTCTCTCCACCAATTTTGGCTCAT	1060
Db	444616	GCGTGGCAACCAACGGGAACAGGAAAACTGCTGTTTTTATTAACCTGGCGCTCAACAT	444555
Qy	1061	ATGATGCATGATGGAAATTACTGCCAGTGGTTTTTAAAGATTGACAGAAACGAGCTGTAAT	1120
Db	444556	TTTATGATATATCCAGCCGTAAACACGAGCCC-----ACCGAGTATTTTG	4445121
Qy	1121	ATGTGACACCAACTCGAAGATTGGTCAACCGATTTATTTGGAAAGCAGAAAAATTTCT	1180
Db	444511	GTAATTAAACCAACCCGTGACCTGCAATGCAAGTGGCTGAAACAAGCGAAGAAATTAGCG	444453E
Qy	1181	TTTGGACTTGTGTAAAGACTCTTGTATATATGSGGGAAACCCAGCTGGGACATTCATT	1240
Db	444451	CAGTTCACCCATTTAAATATTTCCGCAAAATACAGGTGCGTGGGTATCAAAATCACAGGT	4443392
Qy	1241	CGACAAATATGATCAAGGCTGTATATATATATGCTACTCCGGAAGACTGATGATATC	1300
Db	444391	GATGTAATTCAATACCAATCCAAATTTGGTGGTGGCTGACGAGGCCGTTTGTTCATATAC	4443322
Qy	1301	ATBGGCAAGAAAAGATTGGTCTCAACAGATCAAAATCTTAGTTTGGATGAAGCTGAT	1360
Db	444331	ATTAAAGAAAGAAAATTTGATATGTCGCGTTCCGTTGAAATGCTGATTTTGTGTAAGCCGAT	4442727
Qy	1361	CGCATTTTGATATGGGTTTGGTCCAGAAAATGAAGATGTAATTTCTGCCACAGGATG	1420
Db	444271	AGAAATTTGCAATGGGATTTGGGCAAGATGGGAAAAAATTCGACGTGAAACCCGTTGG	4442121
Qy	1421	CCATCCAAAGAAACAGGCCAAACCCCTTATGTTCACTGACACTTTTCCAGAGAAATTCAA	1480

Db 444211 CGAAAA-----CAACCTTGTTGTTTCTGCAACCTTAGAAGAGAGATTATVTC 444164

QY 1481 AGGTTGCGTCGAGAGTTTAAAGCAATATATCTGTTTGTGCTGTGACCAAGTGGT 1540

Db 444163 GTTCGATTTGCGGAGGGTTTATTGAAATGATCTGTGAANAATGCGGAACCAAGTCGC 444104

QY 1541 GGAGCATGTAGAGATGTTCAACA---GACCGTTCTCAAGTTGGCCAGTTCTCAAAAAGA 1597

Db 444103 CGTGAAGAAGAAAAAATCATATCAATAGTATTTATACGACGACACACAAATGAAACAAATC 444044

QY 1598 GAAAGACTGTTGAATAATCTGCGAANAACATAGGGGATGAAANAACATAGCTTTGTTGAA 1657

Db 444043 AAATTGCTGCGCGGTTTATTGAAACTGAAAGTAAGCAACCGCTGGAAATGTGTTATTTCGT 443988

QY 1658 ACTAGAAAAAAGCAGATTTTACTGCAACTTTTCTTGTCAGAAAAAATATCAACTCA 1717

Db 443983 CGTCGTGAAGATGCACTGTAACTTTCCAAAACATGGCTTAAACAGGACATCTGTTCCGA 443922

QY 1718 AGTATCCATGTGATGCGGAACAGAGAGAGCGGAGCAAGCTCTTGAGATATTTTCGTTT 1777

Db 443923 TATTTGAAGGGGAATGSCCAAACTCAAGTAAACAATGCAATTGATTAATTGAATCA 443864

QY 1778 GGAATGCCCCAGTTCTTGTGTTGCTACTTCACTAGTAGTGCACAGGCGCTGATTTGAAAT 1837

Db 443863 GGTATTTGACGCGTATTTGGTTTGCAACAGATGTGCGTGCACGCTGATTTGATTTGACGAT 443804

QY 1838 GTGCACATGTTATCAATTTGATCTTCTTCTTCAACATGATGATATGTTCATCGAATT 1897

Db 443803 GTTAGCCACGTATGATTTTGTGATTTGCTTCCTATATGTCGGATTAATTTCACATCGAATT 443744

QY 1898 GGGCGTACTGTCGCTTGCGGAATCTGGGACAGACAAATTCCTTTTGA 1947

Db 443743 GGAACGTACCGCGGACCTGCGCAAAAAGGACCGGAGCTCTTTTGTGCA 443694

RESULT 14  
 US-09-643-990A-1/c  
 : Sequence 1, Application US/09643990A  
 : Patent No. 6528289  
 :  
 GENERAL INFORMATION:  
 APPLICANT: Robert D. Fleischmann  
 Mark D. Adams  
 Owen White  
 Hamilton O. Smith  
 J. Craig Venter  
 TITLE OF INVENTION: The Nucleotide sequence of  
 the Haemophilus influenzae Rd Genome, Fragments  
 Thereof, and Uses Thereof  
 NUMBER OF SEQUENCES: 1  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Human Genome Sciences, Inc.  
 STREET: 9410 Key West Avenue  
 CITY: Rockville,  
 STATE: MD  
 COUNTRY: USA  
 ZIP: 20850  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: 3 1/2 inch diskette  
 COMPUTER: Dell Pentium  
 OPERATING SYSTEM: MS DOS V6.22  
 SOFTWARE: ASCII Text  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/643,990A  
 FILING DATE: 23-Aug-2000  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/487,429  
 FILING DATE: 1995-06-07  
 APPLICATION NUMBER: 08/426,787  
 FILING DATE: 1995-04-21  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Kenley K. Hoover

```

REGISTRATION NUMBER: 40.302
REFERENCE/DOCKET NUMBER: P8186P1C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-610-5790
TELEFAX: 310-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
Type: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

Query Match      5.6%  Score 125.6;  DB 4;  Length 1830121;
Best Local Similarity 47.3%;  Pred. No. 1.3e-24;
Matches 506;  Conservative 0;  Mismatches 534;  Indels 30;  Gaps 3;

QY      881  TTGAAGAGCTAATCTCTGTCAAGACATGATATACAACTTGCTAAAGCTGTTACT 940
      444736  TTGAACATTCATCTTCTCTGAGCTTTTAAAGCACTTGAAAAAGTTATTC 444677

QY      941  AAGTTCTCTGTCGCAAAATACAGTATCTTATCATCTTCGAGAGCAAGTTGANG 1000
      444676  CGCCCAACGCTATTCMAATGGAAGCCATTCTCGCATGGAAGAGATGATATTA 444617

QY      1001  GCTTGCTCAAAACAGGAGTGTGGGAAGATGCGGCTTTCTCTACCAATTTGGCTCAT 1060
      444616  GGCTCGGACCAACAGGGAAGAGGAAACGCTGCTTTTATTAACCTGGCTACAACT 444557

QY      1061  ATGATGATATGAGATTAATCTGCCAGTCTGTTTAAAGTTCGACGAACGAGTGTAT 1120
      444556  TTATGATATATCCACGCCGTAAACGAGGCC-----ACACCTATTTTG 444512

QY      1121  ATTGTACCACTCACTGAGAAATGCTACCAAGATTATTTGGAGCGAAATTTTCT 1180
      444511  GTATTAACCAACCCGTGACTGCAATGCAAGTGGCTGAACAGCGGAAGATTAGCG 444452

QY      1181  TTGGGACTGTGTAGAGCTGTTGTATATATGAGGGAACCCAGCTGGGACATTCAT 1240
      444451  CAGTTCAACCATTAATATTTGGACATATACAGTGGCGTATCAAAATCACGGT 444392

QY      1241  CGACAAATAGTACAGGCTGTATATATATATGCTTACTCTTGGAAGCTAGATATC 1300
      444391  GATGATTAATCAATCAATCAATCAAGATTGCTGCTAGCGCAAGCGGTTGTTCAT 444332

QY      1301  ATAGCAAAAGAAAGATGCTCTCAACAGATCAAAATTAATTTTGTGATGAGCTGAT 1360
      444331  ATTAAGAGAAAGAAATTTGATGCGGTCGTTGAATGCTGATTTTGAATGAGCGAT 444272

QY      1361  CGCATGTTGATGAGTGGTGTGCTCCAGAAATGAAGAAGTAAATTTCTGCGCAGAA 1420
      444271  AGAATGTTGAATGGAATTTGGCAGATGCGAAAAAATTCAGCTGAACCCGTG 444212

QY      1421  CCATCAAGAGAACGCGCAACCCCTATGTTGACAGCACTTTCCAGAGAAATTCAA 1480
      444211  CGAAAA-----CAAATTTGTTGTTTCTTGCAACCTTAAGAGAGATTAATG 444164

QY      1481  AGGTTGGCTGCAAGTTTTAAAGTCAATTAATCTGTTGTTGCTGTTGCAAGTGGT 1540
      444163  GTGATTTCCGCGGCGCTTATTAAGATGATCCGTGAAGTAAATGATCGGAACCA 444104

QY      1541  GAGACATGTAGAGTGTTCAGCA---GACGCTCTCCAGATTGGCCAGTTCTCAAAA 1597
      444103  CGTGAAGAAAAAATCAATCAATGATATTAACGCAACGCAATGAACCAAAATC 444044

QY      1598  GAAAAGCTGTTGAATTTGCGAAACATAGGAGATGAAGAACTATGTCCTTTGTGA 1657
      444043  AAATTTGCTCGCGCTTTATTAAGAACTGAAGAGTAAACCGTGAATTTGTTATTC 443984

QY      1658  ACTAAGAAAAAGAGATTTTACGCAACTTTCTTTGTGTAAAGAAAAATATCACTAA 1717
      443983  CGTGTGAAGATGACGTGAACCTTTCGAAACATTTGCGTAAACGAGGCAATTCG 443924

```

```

QY      1718  AGTATTCATGTGATCGGGAACAGAGAGGGGAGCAAGCTCTTGAGATTTTGCCTT 1777
      443923  TATTTAAGAGGGGAATGCAACAACTCAAGTAACTGCAATGATTAATTAATCA 443864

QY      1778  GGAAGTCCAGTCTTCTGTTGCTACTTCACTAGTGCAGAGGCTGATATGAAAT 1837
      443863  GGTATTTGAGCGGATTTGTTGCAACAGATGCGTGCACGTTGATGATTAACGAT 443804

QY      1838  GTGCAACATGTTATATTTTGTATTTCTTCTTACATGATGAATATGTCATGAT 1897
      443803  GTAGCCAGCGTAAATTTGATTTGCTTATGTTGGATGATTAATTTGCAATGAA 443744

QY      1898  GGGGCTACTGTCGTTGTGGGAATCTGCGAGCAATTTCTTTTGA 1947
      443743  GGACGTACCGCGCGAGCTGGCAAAAAAGCAAGCAGTCTCTTTGTGCA 443694

Db      443743  GGACGTACCGCGCGAGCTGGCAAAAAAGCAAGCAGTCTCTTTGTGCA 443694

RESULT 15
PCT-US96-05320A-894
; Sequence 894, Application PC/TUS9605320A
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences
; APPLICANT: 9410 Key West Avenue
; APPLICANT: Rockville, MD 20850
; APPLICANT: United States of America
; APPLICANT: Johns Hopkins University
; APPLICANT: 720 Rutland Avenue
; APPLICANT: Baltimore, MD 21205
; APPLICANT: United States of America
; APPLICANT: Mark D. Adams
; APPLICANT: Owen White
; APPLICANT: Hamilton O. Smith
; APPLICANT: J. Craig Venter
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome.
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20003-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/05320A
; FILING DATE: April 122, 1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/476,102
; FILING DATE: June 7, 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,429
; FILING DATE: June 7, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Eric K. Steffe
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488, 014PC01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 894:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1254 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
PCT-US96-05320A-894

```



Query Match 5.6%; Score 123.6; DB:5; Length 1254;  
Best Local Similarity 47.1%; Pred. No. 7.9e-26;  
Matches 495; Conservative 0; Mismatches 539; Indels 18; Gaps 3;

QY 921 TTCTTAAGCTGGTTTACTAAGCTTACTCTGTGTGCAAAAACAGTATTCCTATCAATC 980  
DB 80 TGGCAAAAAGGCTTTGATTTTGTACCCTTAATCCGCTTATCCTTACAGTT 139  
QY 981 TTGCAGAGAGATTTGATGAGCTTGTCTCAACAGGCTGTGGAAGACTGCGGCTTTC 1040  
DB 140 TAAATGACGAGATGTGCGACAGCAAGCTCAAACTGTATCAGCAAGACATGCGCTTTT 199  
QY 1041 TCCTACCAATTTTGGCTCATATGATGATGAGAAATTAACGCCAGTGTTTAAAGAT 1100  
DB 200 TAACGGCTACTTTTCACCTCTTTTAACACCAAGA-----TCTTAATCTTAAT 250  
QY 1101 TGCAGAACCCAGCTGATATTAATGTAGCAACCACTGAGAAATTTGTCAACAGATTAAT 1160  
DB 251 ATCTTACCTCAAGACTTTGATTTTGAACCTTACTGTGAGAAATTAAGCGTACAGATTAGTA 310  
QY 1161 TGGAGCCAGAAAATTTCTTTTGGACTGTGTAAAGACTGTGTATATATATGAGGAGAA 1220  
DB 311 ATGACGACAGATTTCTGCAAAAAGCGATGATTAAGACCGCACTTGCTATGAGTGGG 370  
QY 1221 CCCAGCTGGACATTGAATTCGCAAAATGATCAAGGCTGTAAATATATATGTGCTACTC 1280  
DB 371 ATGTTATGATTAACACTACCAAGCGATGAGCTGCGTCATTTTGATTTGATGATGCA 430  
QY 1281 CTGGAAGACTGATGATATCATATGAGCAAAAAGATTTGTCTCAACAGATCAAAATCT 1340  
DB 431 CGGGCGGAGTCATTTGATTTATGTGAACAAAGGCTAATTTGTTAGTGAATCCAAAGTTG 490  
QY 1341 TAGTTTGGATGAGCTGATCGCATGTGTGATATGAGTTTGTGTCAGAAATGAAGAAAT 1400  
DB 491 TCGTGTAGATGAGCAGATCGAATGTTGATCTTGGGTTTATCCGTGATATTCGTTAT 550  
QY 1401 TAAATTCCTGCCAGGATGCCATCAAGAAACAGGCGCAACCTTATGTCAATGCA 1460  
DB 551 TATTTGGTAA-----ATGCCCGCTCCGCAAGCTGTTAAAGATTTATTTTCAAGCA 604  
QY 1461 CTTTCCAGAGAAATTCAAAGSTTGGCTGACAGAGTTTAAAGTCAATTAATCTGTTTG 1520  
DB 605 CGCTTCTTAATAAGCGCTGATTTAGCATTTGAAGATATGATGAACCTGA---ATA 661  
QY 1521 TTGCTGTGAGCAAGTGGGCTGAGCATGAGATGTTGAGCAGACCGTTCTCCAAAGTTG 1580  
DB 662 TTGAAAATGACAGAACAAAACAGGGCACCGAATTAAGAGAACTTTTATTCAT 721  
QY 1581 GCCAGTTCATAAAGAGAAAAGCTGTTGAATTCGCAAAACATAGGGGATGAAGAA 1640  
DB 722 CTAAATCAGATTAATGGCACTTCTTAACCTTAATGAGAGTGAATGGCTGAACGCT 781  
QY 1641 CTATGCTCTTTGTTGAACTTAAGAAAAGCAGATTTTACGCACTTTCTTTGTCAAG 1700  
DB 782 GTATTTGATTTGGAAATACGAACATGTTGTGAACAAATTTGGGCTATTTTGGCGGCTG 841  
QY 1701 AAAAATATCAATCAAGATATCATGATCGGAGACAGAGAGCGGAGCAAGCTC 1760  
DB 842 ATGGGATGCTGTGCGTTTACTGACTGGGATGTAGCAGAAAAACGTTATCGTTAT 901  
QY 1761 TTGGAATTTTCCCTTTGAAAGTGGCCAGTTCTTGTCTTACTGATAGCTGCCAAG 1820  
DB 902 TAAACAAATTTACTGATGATTTGATTAATTTTGTGCAACAGATGCGCTGCTG 961  
QY 1821 GCGTGGATTTGAAAAATGTGAACATGTTATCAATTTGATCTTCTTACCATTTGATG 1880  
DB 962 GCTTGCATATTTCTGATGTAGCAGCATTTTCAATTAATGATTTACCCGATGATCGAAG 1021  
QY 1881 AATATGTTCATCGAATTTGGCGGCTGCTGCTGTTGTGGAATACCTGGCAGAGCAATTTCT 1940  
DB 1022 ATATATGTTCAACGAATTTGGCGGCTGCTGAGCAGACAGGGAAGTGTGTTGATTAATG 1081  
QY 1941 TTTTGTATCTGATGCAATACCAATTGCA 1972

DB 1082 TCGCTGTGAAGATATGCGATGATTAACA 1113

Search completed: July 27, 2004, 10:52:53  
Job time : 189 secs



XX Novel vasa polynucleotides useful in the diagnosis or treatment of  
PT conditions characterized by aberrant expression and/or presence of mutant  
PT forms of vasa polynucleotides or polypeptides.

XX Claim 1: Page 52-53; 66pp; English.

CC The present sequence is human vasa cDNA that has germ cell specific  
CC expression and is believed to play a determinative role in gonad  
CC development. Germ cells are specialised to produce haploid gametes in  
CC multicellular organisms. Vasa is useful in the diagnosis or treatment of  
CC conditions characterised by its aberrant expression and/or the presence  
CC of its mutant forms. The conditions include cancers such as biliary  
CC tract, brain, breast, colon, ovarian, pancreas, prostate, colorectal,  
CC oral, liver, lung, skin, basocellular, testis, renal, thyroid, cervical,  
CC endometrial, oesophageal and gastric, lymphomas, melanomas,  
CC glioblastomas, neuroblastomas, medulloblastomas, choriocarcinoma,  
CC squamous cell carcinoma, haematological neoplasms, acute lymphocytic and  
CC myelogenous leukaemia, multiple myeloma, Acquired immune deficiency  
CC syndrome (AIDS) associated leukaemias, intraepithelial neoplasms, Bowen's  
CC disease, Paget's disease, sarcomas such as leiomyosarcoma,  
CC rhabdomyosarcoma, liposarcoma, fibrosarcoma, Kaposi's sarcoma and  
CC osteosarcoma, tumours such as testicular tumour (eg. seminoma), ovarian  
CC tumour (eg. dysgerminoma or teratoma) and tumour of an extragonadal  
CC tissue (eg. mediastinal or an intracranial tumour)

XX Sequence 2224 BP; 678 A; 385 C; 550 G; 611 T; 0 U; 0 Other;

Query Match 100.0%; Score 2224; DB 5; Length 2224;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2224; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ACTTGAAGTCACCATGGGGGATGAAGTTGGGAAAGCAGAAATCAACCTCATATGCTTC 60
DB 1 ACTTGAAGTCACCATGGGGGATGAAGTTGGGAAAGCAGAAATCAACCTCATATGCTTC 60
QY 61 CTATGTTCCCATATTTGAGAGGATAGTATTTCTGAGAGAAAATGAGACATTTTACAG 120
DB 61 CTATGTTCCCATATTTGAGAGGATAGTATTTCTGAGAGAAAATGAGACATTTTACAG 120
QY 121 GACTCCAGCTTTCATCAGAAATGATGACCTTCTCGAAGATCATTTTCATGAA 180
DB 121 GACTCCAGCTTTCATCAGAAATGATGACCTTCTCGAAGATCATTTTCATGAA 180
QY 181 AAGTGAATTTGCTCTGGGGGGAATTTTGGAAAACAGAGATGCTGGTATATTAAGCG 240
DB 181 AAGTGAATTTGCTCTGGGGGGAATTTTGGAAAACAGAGATGCTGGTATATTAAGCG 240
QY 241 AGATTAATATCATCCAAATGGGTGTTTGGAGTTGAAAAGAGTTTGGAAACAGAGCTT 300
DB 241 AGATTAATATCATCCAAATGGGTGTTTGGAGTTGAAAAGAGTTTGGAAACAGAGCTT 300
QY 301 TTCAAAACAGAGGTTGAGAGATGATAGTCTGCTTCTGAGAGAGCTCTAATATGA 360
DB 301 TTCAAAACAGAGGTTGAGAGATGATAGTCTGCTTCTGAGAGAGCTCTAATATGA 360
QY 361 CTGCGAAGATTAATCCACAACGGAACAGAGGTTTTCAGAGAGGCGCTATCGAGATGG 420
DB 361 CTGCGAAGATTAATCCACAACGGAACAGAGGTTTTCAGAGAGGCGCTATCGAGATGG 420
QY 421 AATTAATTCAGAAAGCTTCAGGGCCATACAGAAAGGTGAGAGATGTTTCCGAGTTG 480
DB 421 AATTAATTCAGAAAGCTTCAGGGCCATACAGAAAGGTGAGAGATGTTTCCGAGTTG 480
QY 481 CCGTGAAGAGATTGGTCTAGAAAGTCCAATATATGACTTAGACCCAGAGGAATGATGCA 540
DB 481 CCGTGAAGAGATTGGTCTAGAAAGTCCAATATATGACTTAGACCCAGAGGAATGATGCA 540
QY 541 GGGCATGTGGTGGCTTTTGGTTTCAAGAACCAAGATTTAATGAGCAACAGGTAATGCTGA 600
DB 541 GGGCATGTGGTGGCTTTTGGTTTCAAGAACCAAGATTTAATGAGCAACAGGTAATGCTGA 600
QY 601 TACTTCTCAAGCAGAGAGTGCAGTGAAGTGAACGAGGTGGTTCAAAAGTTTAAATGA 660
```

```
DB 601 TACTTCTCAAGCAGAGAGTGCAGTGAAGTGAACGAGGTGGTTCAAAAGTTTAAATGA 660
QY 661 AGAATTAATTAACAGGCTCTGAAAAGATTTCTTGAAGTCAGAACAGAGAGAGAGAAAG 720
DB 661 AGAATTAATTAACAGGCTCTGAAAAGATTTCTTGAAGTCAGAACAGAGAGAGAGAAAG 720
QY 721 TAGTATATCTCAAGAGCAAAAGTGAACATATACCCCTCTCACCTGAGATGAGGA 780
DB 721 TAGTATATCTCAAGAGCAAAAGTGAACATATATACCCCTCTCACCTGAGATGAGGA 780
QY 781 CTCCATCTTTGACATTTTCAAGACGATTAATCTTGACAAATACGACTATTTCTGT 840
DB 781 CTCCATCTTTGACATTTTCAAGACGATTAATCTTGACAAATACGACTATTTCTGT 840
QY 841 GGAATGCTGACATGATGACACACGACATTTCTGATTAAGAAGTATCTCTG 900
DB 841 GGAATGCTGACATGATGACACACGACATTTCTGATTAAGAAGTATCTCTG 900
QY 901 TCAAGACTGAATTAACAACATTTGCTAAAGCTGTTATATGACTTACCTCTGCAAAA 960
DB 901 TCAAGACTGAATTAACAACATTTGCTAAAGCTGTTATATGACTTACCTCTGCAAAA 960
QY 961 ATACAGTATTCCTATCAATCTTGACAGACGATTTGATGGCTTGCTCAACAGGCTC 1020
DB 961 ATACAGTATTCCTATCAATCTTGACAGACGATTTGATGGCTTGCTCAACAGGCTC 1020
QY 1021 TGGGAAGACTGGGCTTTTCTCTACCAATTTTGGCTCATATGATGATGAGATTAAC 1080
DB 1021 TGGGAAGACTGGGCTTTTCTCTACCAATTTTGGCTCATATGATGATGAGATTAAC 1080
QY 1081 TGCCAGTCGTTTAAAGATTGACAGAACCAAGATGATATTTGTAGACCAACTGAGA 1140
DB 1081 TGCCAGTCGTTTAAAGATTGACAGAACCAAGATGATATTTGTAGACCAACTGAGA 1140
QY 1141 ATTGCTCAACAGATTTATTTTGAAGCCAAATAATTTCTTTGGGACTGTGTAAGAGC 1200
DB 1141 ATTGCTCAACAGATTTATTTTGAAGCCAAATAATTTCTTTGGGACTGTGTAAGAGC 1200
QY 1201 TGTGTTATATATGCGGGAACCCAGCTGGGACATTCATTTGCAAAATAGTACAGGCTG 1260
DB 1201 TGTGTTATATATGCGGGAACCCAGCTGGGACATTCATTTGCAAAATAGTACAGGCTG 1260
QY 1261 TAAATATATATATGCTCTCTGGAAGACTGATGGAATATGAGGCAAAAGAAATTTGG 1320
DB 1261 TAAATATATATATGCTCTCTGGAAGACTGATGGAATATGAGGCAAAAGAAATTTGG 1320
QY 1321 TCTCAACAGATCAAAATATAGTTTGAATGAGAGCTGATCGCATGTTGATAGGCTT 1380
DB 1321 TCTCAACAGATCAAAATATAGTTTGAATGAGAGCTGATCGCATGTTGATAGGCTT 1380
QY 1381 TGGTCCAGAAATGAAGAGTTAATTTCTTCCAGAGAAATGCCATCAAAAGAACAGGCCA 1440
DB 1381 TGGTCCAGAAATGAAGAGTTAATTTCTTCCAGAGAAATGCCATCAAAAGAACAGGCCA 1440
QY 1441 AACCTTATGATGATGCAACTTTTCCAGAGAAATTCAAAGGTTGGCTGACAGATTTT 1500
DB 1441 AACCTTATGATGATGCAACTTTTCCAGAGAAATTCAAAGGTTGGCTGACAGATTTT 1500
QY 1501 AAAGTCAAAATTAATCTGTTTGTCTGTTGGAACAAGGGGTGAGCATGTAGAGATGTTCA 1560
DB 1501 AAAGTCAAAATTAATCTGTTTGTCTGTTGGAACAAGGGGTGAGCATGTAGAGATGTTCA 1560
QY 1561 GCAGACCGTCTCAAGTTGGCAGTTTCAAAAAGAAAAGAGAGCTGTTGAATTTCTGGG 1620
DB 1561 GCAGACCGTCTCAAGTTGGCAGTTTCAAAAAGAAAAGAGAGCTGTTGAATTTCTGGG 1620
QY 1621 AAACATAGGGAGTGAAGAACTATGATCTTTGTTGAACCTAAGAAAAGAGAGATTTTAC 1680
DB 1621 AAACATAGGGAGTGAAGAACTATGATCTTTGTTGAACCTAAGAAAAGAGAGATTTTAC 1680
QY 1681 TGCACCTTTCTTGTGCAAGAAAATATCAATCAAGATTCATGATGCTGGGAAAC 1740
```

Db	1861	TGCAACTTTTCTTTGTGCAGAAAAAAATATCAACTACAAGATCATGGTGTGGGAGACA	1740
QY	1741	GAGGAGGGGGGAGAGAGCTCTTGAGATTTTCGCTTTGGAAAGGCCAGTCTTGTTGCG	1800
Db	1741	GAGGAGGGGGGAGAGAGCTCTTGAGATTTTCGCTTTGGAAAGGCCAGTCTTGTTGCG	1800
QY	1801	TACTTCAGTAGCTGCGCAGAGGGCTGGATATTGAAAAATGTGCACACATGTTATCAATTTTGA	1860
Db	1801	TACTTCAGTAGCTGCGCAGAGGGCTGGATATTGAAAAATGTGCACACATGTTATCAATTTTGA	1860
QY	1861	TCTTTCCTTCACATTTGAATGAAATANGTCATACGAAATTTGGGGGTCTGGTCTGTTGGGGA	1920
Db	1861	TCTTTCCTTCACATTTGAATGAAATANGTCATACGAAATTTGGGGGTCTGGTCTGTTGGGA	1920
QY	1921	TACTGGCAGAGCAATTTCTTTTTTGGATCTTGAATCGGATACCATTTTAGCACAGCCTCT	1980
Db	1921	TACTGGCAGAGCAATTTCTTTTTTGGATCTTGAATCGGATACCATTTTAGCACAGCCTCT	1980
QY	1991	AGTAAATATTTGACAGATGCTGCAACAGAGATGTTCTGATCGTTGGAAAGAAATTCCTT	2040
Db	1991	AGTAAATATTTGACAGATGCTGCAACAGAGATGTTCTGATCGTTGGAAAGAAATTCCTT	2040
QY	2041	TAGTACATACATTTCTTGCTTCAGTGTAGTACAAAGAGAAAACGTGTTTGATCATGTGA	2100
Db	2041	TAGTACATACATTTCTTGCTTCAGTGTAGTACAAAGAGAAAACGTGTTTGATCATGTGA	2100
QY	2101	TACCGAAGAGGGCAGAGACCTTTGAAACAACAGCTGGGTTCCTTTACAGAGCTCCCAA	2160
Db	2101	TACCGAAGAGGGCAGAGACCTTTGAAACAACAGCTGGGTTCCTTTACAGAGCTCCCAA	2160
QY	2161	TCCAGTAGATGATGAGTCAATGGGATTTAAAGCCAAAAACATCCTTCAAGTCTGTGTTTGA	2220
Db	2161	TCCAGTAGATGATGAGTCAATGGGATTTAAAGCCAAAAACATCCTTCAAGTCTGTGTTTGA	2220
QY	2221	TGCA 2224	
Db	2221	TGCA 2224	

XX	AD863498	standard; cDNA; 1984 BP.
XX	AD863498;	
XX	04-DEC-2003	(first entry)
XX	Human CDNA encoding clone TESTI20227380.	
XX	Human; ss; gene; pharmaceutical; diagnostic; gene therapy;	
KW	tissue regeneration; cell regeneration; membrane protein;	
KW	signal transduction-related protein; transcription-related protein;	
KW	osteoporosis; neurological disease; cancer; tumour.	
XX	Homo sapiens.	
OS		
XX	Key	Location/Qualifiers
FH	CDS	20..1747
FT		/tag= a
FT		/product= "Clone TESTI20227380 protein"
FN	EPI308459-A2.	
XX		
PD	07-MAY-2003.	
XX		
Pf	28-MAR-2002; 2002EP-00007401.	
XX		
PR	05-NOV-2001; 2001JP-00379298.	
PR	25-JAN-2002; 2002US-00350978.	
XX		
PA	(HELI-) HELIX RES. INST.	
PA	(REAS-) RES ASSOC BIOTECHNOLOGY.	
XX		

PI Isgaard T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;  
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Rife R, Tamechika I;  
PI Seki N, Yoshikawa T, Otsuka M, Nagahara K, Maeno Y;  
XX WPI; 2003-450961/43.  
DR P-PSDB; ADB65468.  
XX  
XX New polynucleotides and polypeptides, useful for developing a diagnostic  
XX marker or medicines for regulation of their expression and activity, or  
PT as targets of gene therapy.  
PT

Claim 1; Page; 222pp; English.

The invention discloses a polynucleotide comprising a sequence selected from 1970 fully defined nucleotide sequences which encode novel polypeptides. Also claimed is a polypeptide encoded by the polynucleotide or its partial peptide, an antibody binding to the polypeptide or polynucleotide, immunologically assaying the polypeptide or peptide of the polynucleotide by contacting the polypeptide or peptide with the antibody of the encoded protein, and observing the binding between the two, a transformant carrying the polynucleotide in an expressible manner and an antisense polynucleotide. The oligonucleotide is useful as a primer for synthesising the polynucleotide, or as a probe for detecting the polynucleotide. The polynucleotides and encoded proteins are useful as pharmaceutical agents and many disease-related genes may be included in them, for developing a diagnostic marker or medicines for regulation of their expression and activity, or as targets of gene therapy. The genes are involved in tissue and/or cell regeneration. Membrane proteins, signal transduction-related proteins, transcription-related proteins, disease-related proteins and genes encoding them can be used as indicators for diseases (e.g. osteoporosis, neurological diseases, cancer, tumours. The cDNA may be used to regulate the activity or expression of the encoded protein to treat diseases. The sequence presented is a cDNA of the invention. Note: Some of the sequence data for this patent is not represented in the printed specification, but is based on sequence information supplied by the European Patent Office.

Query Match	72.0%;	Score 1601;	DB 9;	Length 1984;
Best local Similarity	96.9%;	Pred No. 0;		
Matches 1662;	Conservative	0;	Mismatches 5;	Indels 48;
				Gaps 1

Qy	510	ATAATGACTTAACCCAGACGAATGATGACAGCATGAGTGGCCCTTTGGTCTAGAA	569
Db	117	ATAATGACTTAACCCAGACGAATGATGACAGCATGAGTGGCCCTTTGGTCTAGAA	176
Qy	570	GACCAGTATTAAGTGGACAGTAATGGTACTTCTTCAAGCAGAACTGGCAGTGGAA	629
Db	177	GACCAGTATTAAGTGGACAGTAATGGTACTTCTTCAAGCAGAACTGGCAGTGGAA	236
Qy	630	GTGAACGAGTGGTTACAAAGTTTAATGAAGAAATTAACAGGCTTGGAAAGATTT	689
Db	237	GTGAACGAG-----ATT 248	
Qy	690	CTTGGAGTCTAGAGCAGAAAGGAGAAAGTATCTCAAGACCAAAAGTACCT	749
Db	249	CTTGGAAAGTCAGAAAGCAGAAAGGAGAAAGTATCTCGAGGACCAAAAGTACCT	308
Qy	750	AACATACCCCTCTCCACCTGAGAGATGAGGATCCATCTTTGCACATTATCAGACAGCA	809
Db	309	AACATACCCCTCTCCACCTGAGAGATGAGGATCCATCTTTGCACATTATCAGACAGCA	368
Qy	810	TAACTTCGACAAATTCGACACTATTCTTGTGGAAGTCTGTGACATGATGACACACAG	869
Db	369	TAACTTCGACAAATTCGACACTATTCTTGTGGAAGTCTGTGACATGATGACACACAG	428
Qy	870	CATTTCTGACTTTTGAAGAGTAACTCTGTGACACATGAATAACACATTGCTAAAG	929
Db	429	CATTTCTGACTTTTGAAGAGTAACTCTGTGACACATGAATAACACATTGCTAAAG	488
Qy	930	CTGTTATCTAAGCTTACTCTCTGTGCAAAAATACAGTATTCCTATCATCTTGCAGGAC	989



700 **ATTCTTCTCCAAAGTGTCTGGA**CTATGATGCAACCAAGCAATTCTGACTTTGAAGAAAGCT 759

14-AUG-2000; 2000US-0225270P.

14-AUG-2000; 2000HS-0225270E.

PR	14-AUG-2000	2000US-0225447P	PR	08-NOV-2000	2000US-0246532P
PR	14-AUG-2000	2000US-0225157P	PR	08-NOV-2000	2000US-0246609P
PR	14-AUG-2000	2000US-0225158P	PR	08-NOV-2000	2000US-0246610P
PR	14-AUG-2000	2000US-0225159P	PR	08-NOV-2000	2000US-0246611P
PR	22-AUG-2000	2000US-0226279P	PR	08-NOV-2000	2000US-0246613P
PR	22-AUG-2000	2000US-0226681P	PR	17-NOV-2000	2000US-0249207P
PR	22-AUG-2000	2000US-0226688P	PR	17-NOV-2000	2000US-0249208P
PR	23-AUG-2000	2000US-0227182P	PR	17-NOV-2000	2000US-0249209P
PR	30-AUG-2000	2000US-0227709P	PR	17-NOV-2000	2000US-0249210P
PR	01-SEP-2000	2000US-0228924P	PR	17-NOV-2000	2000US-0249211P
PR	01-SEP-2000	2000US-0229343P	PR	17-NOV-2000	2000US-0249212P
PR	01-SEP-2000	2000US-0229344P	PR	17-NOV-2000	2000US-0249213P
PR	05-SEP-2000	2000US-0229309P	PR	17-NOV-2000	2000US-0249214P
PR	06-SEP-2000	2000US-0229513P	PR	17-NOV-2000	2000US-0249215P
PR	06-SEP-2000	2000US-0230437P	PR	17-NOV-2000	2000US-0249216P
PR	08-SEP-2000	2000US-0231242P	PR	17-NOV-2000	2000US-0249217P
PR	08-SEP-2000	2000US-0231243P	PR	17-NOV-2000	2000US-0249218P
PR	08-SEP-2000	2000US-0231244P	PR	17-NOV-2000	2000US-0249244P
PR	08-SEP-2000	2000US-0231411P	PR	17-NOV-2000	2000US-0249245P
PR	08-SEP-2000	2000US-0231414P	PR	17-NOV-2000	2000US-0249246P
PR	08-SEP-2000	2000US-0232080P	PR	17-NOV-2000	2000US-0249247P
PR	12-SEP-2000	2000US-0231968P	PR	17-NOV-2000	2000US-0249297P
PR	14-SEP-2000	2000US-0233397P	PR	01-DEC-2000	2000US-0250160P
PR	14-SEP-2000	2000US-0233398P	PR	01-DEC-2000	2000US-0250391P
PR	14-SEP-2000	2000US-0233399P	PR	05-DEC-2000	2000US-0251030P
PR	14-SEP-2000	2000US-0233400P	PR	05-DEC-2000	2000US-0251031P
PR	14-SEP-2000	2000US-0233401P	PR	05-DEC-2000	2000US-0251988P
PR	14-SEP-2000	2000US-0233063P	PR	06-DEC-2000	2000US-0256719P
PR	14-SEP-2000	2000US-0233064P	PR	08-DEC-2000	2000US-0251479P
PR	14-SEP-2000	2000US-0233065P	PR	08-DEC-2000	2000US-0251856P
PR	21-SEP-2000	2000US-0233223P	PR	08-DEC-2000	2000US-0251863P
PR	21-SEP-2000	2000US-0233227P	PR	08-DEC-2000	2000US-0251869P
PR	25-SEP-2000	2000US-0234997P	PR	11-DEC-2000	2000US-0251909P
PR	25-SEP-2000	2000US-0234998P	PR	05-JAN-2001	2001US-0259678P
PR	26-SEP-2000	2000US-0235484P	XX		
PR	27-SEP-2000	2000US-0235834P	XX		
PR	27-SEP-2000	2000US-0235836P	XX		
PR	29-SEP-2000	2000US-0236327P	XX		
PR	29-SEP-2000	2000US-0236367P	XX		
PR	29-SEP-2000	2000US-0236368P	XX		
PR	29-SEP-2000	2000US-0236369P	XX		
PR	29-SEP-2000	2000US-0236370P	XX		
PR	02-OCT-2000	2000US-0237037P	XX		
PR	02-OCT-2000	2000US-0237038P	XX		
PR	02-OCT-2000	2000US-0237039P	XX		
PR	02-OCT-2000	2000US-0237040P	XX		
PR	13-OCT-2000	2000US-0239935P	XX		
PR	13-OCT-2000	2000US-0239937P	XX		
PR	20-OCT-2000	2000US-0241211P	XX		
PR	20-OCT-2000	2000US-0241212P	XX		
PR	20-OCT-2000	2000US-0241785P	XX		
PR	20-OCT-2000	2000US-0241786P	XX		
PR	20-OCT-2000	2000US-0241787P	XX</		



Best Local Similarity 99.8%; Pred. No. 3.2e-110; Matches 443; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 1781 AAGTGGCCAGTTCTTGTGCTTACTGACAGTCCAGAGGGCTGCATTTGAAATGTG 1840
    |||
Db 1 AAGTGGCCAGTTCTTGTGCTTACTGACAGTCCAGAGGGCTGCATTTGAAATGTG 60
QY 1841 CAACATGTTATCAATTTGATCTCTCTTCAACATGATGAATGTTATGAAATGGG 1900
    |||
Db 61 CAACATGTTATCAATTTGATCTCTCTTCAACATGATGAATGTTATGAAATGGG 120
QY 1901 CGTACTGCTGCTGTGGGAATCTGACAGACAAATTCCTTTTGTGATCTGATGAT 1960
    |||
Db 121 CGTACTGCTGCTGTGGGAATCTGACAGACAAATTCCTTTTGTGATCTGATGAT 180
QY 1961 AACCATTTAGCACAGCCTCTAGTAAAGTATTTGACAGATGCTCAACAGATGTTCTGCA 2020
    |||
Db 181 AACCATTTAGCACAGCCTCTAGTAAAGTATTTGACAGATGCTCAACAGATGTTCTGCA 240
QY 2021 TGGTTGGAGAATTTGCTTTAGTACATACATTCCTGCTTCACTGCTGATGACAGAGA 2080
    |||
Db 241 TGGTTGGAGAATTTGCTTTAGTACATACATTCCTGCTTCACTGCTGATGACAGAGA 300
QY 2081 AACGTGTTTCATCAGTTGATACACAGAAAGGCAAGACCTTTGAACACAGCTGGGTTT 2140
    |||
Db 301 AACGTGTTTCATCAGTTGATACACAGAAAGGCAAGACCTTTGAACACAGCTGGGTTT 360
QY 2141 TCTTCTTCACAGAGTCTCCCAATCCAGTATGATGATGATGATGATGATGATGATGATG 2200
    |||
Db 361 TCTTCTTCACAGAGTCTCCCAATCCAGTATGATGATGATGATGATGATGATGATGATG 420
QY 2201 CTTCAAGCTGTGTGTTTATGCA 2224
    |||
Db 421 CTTCAAGCTGTGTGTTTATGCA 444
```

## RESULT 5

ABZ20094  
ID ABZ20094 standard; cDNA; 667 BP.

AC ABZ20094;

DT 23-JAN-2003 (first entry)

DE Group III cDNA cancer related clone SEQ ID NO:2520.

XX Human; Cancer; tumour; therapy; diagnosis; CT antigen; CP antigen;  
XX immune response; virology; immunology; microbiology; molecular biology;  
KW recombinant DNA technology; gene; ss.

OS Homo sapiens.

PN MO200278516-A2.

PD 10-OCT-2002.

PF 28-MAR-2002; 2002WO-US010421.

PR 30-MAR-2001; 2001US-0280255P.

PR 28-AUG-2001; 2001US-0315563P.

PR 09-JAN-2002; 2002US-0347313P.

PA (CORI-) CORIXA CORP.

PI Wang T, Wang S, Bangur CS, Gaiger A;

DR WPI; 2003-058387/05.

XX New immunogenic polynucleotides or polypeptides useful for diagnosing,  
PT preventing and treating cancer expressing CT or CP mRNA antigens, and in  
PT virology, immunology, microbiology, molecular biology and recombinant DNA  
PT techniques.

PS Claim 1; SEQ ID NO 2520; 207pp; English.

XX AB017575 to AB020506 represent isolated polynucleotide (I) sequences, and  
CC AB017575 to AB020506 represent protein (II) sequences, from the present  
CC invention. (I) and (II) have cytostatic activity and can be used in gene  
CC therapy and vaccines. (I), (II), antibodies and compositions from the  
CC present invention are useful for diagnosing, preventing and treating  
CC cancer, which expresses CT or CP mRNA antigens. They are useful for  
CC stimulating immune response. They can also be useful in virology,  
CC immunology, microbiology, molecular biology and recombinant DNA  
CC techniques. N.B. The sequence data for this patent did not form part of  
CC the printed specification, but was obtained in electronic format directly  
CC from Wipo at [http://wipo.int/pub/publ/published\\_poc\\_sequences](http://wipo.int/pub/publ/published_poc_sequences)

XX Sequence 667 BP; 215 A; 146 C; 155 G; 150 T; 0 U; 1 Other;

Query Match 19.6%; Score 435.2; DB 7; Length 667;

Best Local Similarity 99.3%; Pred. No. 3e-108; Matches 437; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
QY 655 AATGAGAGATTAATTAACAGGCTCTGAAAGATTTCTGAAAGTCAAGACAGAAAGG 714
    |||
Db 1 AATGAGAGATTAATTAACAGGCTCTGAAAGATTTCTGAAAGTCAAGACAGAAAGG 60
QY 715 AGAAGTATGATATCTCAAGACCAAAAGTACCTACCTCCCTCCACCTGAGGA 774
    |||
Db 61 AGAAGTATGATATCTCAAGACCAAAAGTACCTACCTCCCTCCACCTGAGGA 120
QY 775 TGAGACTCCATCTTTGACATTTATCAGACAGGATTAACCTTGCACAAATAGACACTAT 834
    |||
Db 121 TGAGACTCCATCTTTGACATTTATCAGACAGGATTAACCTTGCACAAATAGACACTAT 180
QY 835 TCTTGTGAGAGTGTCTGACATGATGACACACAGCAATTCGATTTTGAAGAGCTAA 894
    |||
Db 181 TCTTGTGAGAGTGTCTGACATGATGACACACAGCAATTCGATTTTGAAGAGCTAA 240
QY 895 TCTCTGTGACACATGATTAACACATTCGTAAGCTGGTATTAATTAAGCTTACTCTGT 954
    |||
Db 241 TCTCTGTGACACATGATTAACACATTCGTAAGCTGGTATTAATTAAGCTTACTCTGT 300
QY 955 GCAAAATACAGTATTCCTATCATCTTGACAGACGAGATTTGATGCTTGTCTCAAC 1014
    |||
Db 301 GCAAAATACAGTATTCCTATCATCTTGACAGACGAGATTTGATGCTTGTCTCAAC 360
QY 1015 AGGCTGTGGAGAGACCTGGGCTTTCTCTACCAATTTTGGCTCATGATGATGATG 1074
    |||
Db 361 AGGCTGTGGAGAGACCTGGGCTTTCTCTACCAATTTTGGCTCATGATGATGATG 420
QY 1075 AATTAAGTCCAGTCTTTTA 1094
    |||
Db 421 AATTAAGTCCAGTCTTTTA 440
```

## RESULT 6

ABX63480  
ID ABX63480 standard; cDNA; 2451 BP.

AC ABX63480;

DT 26-FEB-2003 (first entry)

DE Human cDNA #480 differentially expressed in activated vascular tissue.

XX Human; gene; ss; vascular tissue; cytosolic; atherosclerosis; cardiac;  
KW hypotensive; antidiabetic; gynaecological; vasectomized; cerebroprotective;  
KW gene therapy; vascular disease; cancer; coronary; artery disease;  
KW hypertension; diabetes; pre-eclampsia; restenosis;  
KW :ischemia-reperfusion injury; stroke.

OS Homo sapiens.

PN US2002137081-A1.

PD 26-SEP-2002.  
 XX 08-JAN-2002; 2002US-00044090.  
 XX 28-JUN-2000; 2000US-0222469P.  
 PR 08-JAN-2001; 2001US-0260483P.  
 XX (BAND/) BANDMAN O.  
 PA Bandman O;  
 PI WPI; 2003-110597/10.  
 DR  
 XX  
 XX Combination for diagnosing, staging, treating, or monitoring the  
 PT progression of treatment of a vascular disease, e.g. atherosclerosis,  
 PT comprising several cDNAs that are differentially expressed in activated  
 PT vascular tissue.  
 PS  
 XX Claim 1; Page; 18pp; English.  
 CC This invention relates to a combination comprising several cDNAs that are  
 CC differentially expressed in activated vascular tissue. The invention also  
 CC discloses a high throughput method for detecting differentially expressed  
 CC cDNAs in a sample. The cDNAs of the invention may have  
 CC antiatherosclerotic; cyostatic; cardiac; hypotensive; antidiabetic;  
 CC synaescological; vasotropic and cerebroprotective activities and may be  
 CC used in gene therapy. The cDNAs of the invention may be used in a high-  
 CC throughput methods for detecting differential expression of one or more  
 CC cDNAs in a sample, or screening several molecules or compounds to  
 CC identify a molecule or compound that specifically binds a cDNA of the  
 CC invention. A protein encoded by the cDNA may be used to screen several  
 CC molecules or compounds to identify a ligand that specifically binds to  
 CC the protein, or to produce or purify a ligand that specifically binds to  
 CC be used to detect a protein in a sample or purify a natural or  
 CC recombinant protein from a sample. The nucleotides may be useful for  
 CC diagnosing, staging, treating, or monitoring the progression of treatment  
 CC of a vascular disease, e.g. atherosclerosis, cancer, coronary artery  
 CC disease, hypertension, diabetes, pre-eclampsia, ischemia-reperfusion  
 CC injury, restenosis, or stroke. The cDNAs can also be used for large-scale  
 CC genetic or gene expression analysis of several new nucleic acid  
 CC molecules. Antibodies to the proteins encoded by the cDNAs are useful for  
 CC diagnosing pre-pathologic disorders, and chronic or acute diseases  
 CC associated with abnormalities in the expression, amount or distribution  
 CC of the protein. The present sequence represents a cDNA of the invention  
 CC that is differentially expressed in activated vascular tissue. Note: The  
 CC sequence data for this patent did not form part of the specification, but  
 CC was obtained in electronic format directly from USPTO at  
 CC http://seqdata.uspto.gov/sequence.html?docid=20020137081  
 XX  
 XX  
 XX Sequence 2451 BP; 737 A; 436 C; 613 G; 665 T; 0 U; 0 Other;  
 SQ  
 Query Match 17.4%; Score 386.4; DB 7; Length 2451;  
 Best Local Similarity 58.7%; Pred. No. 1.2e-94;  
 Matches 763; Conservative 0; Mismatches 501; Indels 36; Gaps 4;

Db 743 TTTTACCACTACTGAGTACAGATATATACAGATGCTCCAGGAAAGCTTTGAAGCTGTGA 802  
 Qy 1074 GAATTAACGCACTGCTGTTTAAAGATTGACGAGAACGAGTGTATTAATGTACACCA 1133  
 Db 803 AGGTAAATGGAAGTATGGGCGCCGCAACATATCCATATCTGTTTACGCCCA 862  
 Qy 1134 CTGAGAAATGGTCAACAGATTTATTTGGAAGCCAGAAATTTTCTTTTGGACTGTG 1193  
 Db 863 CAAAGAAATTTGGCTGTACAGATCTATGAGAGAGCCGAAATTTTCCATCCGATCTAG 922  
 Qy 1194 TAAGAGCTGTTTATATATGAGGAGAAACCACTGGGACATTCATTTGACAAATATGAC 1253  
 Db 923 TTGCTCTTGTGTAGTTTATGATGCTGTGTATATGTCAGCAATTCGGGACTTAGAC 982  
 Qy 1254 AAGGTGTAAATATATATATGATGCTGTCTCTGGAACATGATGATATATAGGCAAGAA 1313  
 Db 983 GTGATGCCACTTGTATGATACCACTCAGAGCTCTAGGATATATGAAAGAGAA 1042  
 Qy 1314 AGATTGCTTCCAAACAGATCAATATCTAGTTTGGATGAAAGTATGATGATGATGATA 1373  
 Db 1043 AGATTGATTAAGCTTCTGCAAGTACTAGTGTGATGAAAGTATGATGATGATGATA 1102  
 Qy 1374 TGGGTTTGGTCCGAAATGAAAGATTAATTTCTTCCAGAGATCCATCAAGAGAAC 1433  
 Db 1103 TGGATTTGAACCTCAGATACGCTGATATGTTGAAACAAATATATGCTCACAAGGCG 1162  
 Qy 1434 AGCCCAACCCCTATGTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1493  
 Db 1163 TTGCTACACAT 1222  
 Qy 1494 AGTTTAAAGTCAAT 1553  
 Db 1223 ACTTTTGGATGAA--TATATCTTTTGGCTGATGAGAGATGAGGCTTACTCTGAGA 1279  
 Qy 1554 ATGTTGACGAGACCGCTTCCAGTGTGATGATGATGATGATGATGATGATGATGATGAT 1613  
 Db 1280 ACATCAACAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1339  
 Qy 1614 TTCTGCAAAACATAGG--GATGAAAGAACTATGCTTTTGTGAATCAAGAAAGAG 1670  
 Db 1340 TTTTATGATGACAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1399  
 Qy 1671 CAGATTTTACTGCACTTTCTTTTCCAGAAATATCACTAATCAAGTATCATGATG 1730  
 Db 1400 CAGATTTCTGAGAGATTTCTTATACCAAGAAATATCTGTACTATATATTCATGAG 1459  
 Qy 1731 ATCGGAAACAGAGAGACCGGAGCAAGCTTTGAGATTTTGGCTTTGAAAGTCCAG 1790  
 Db 1460 ACCGTCAGAGAGATGAG 1519  
 Qy 1791 TTCTTGTGCTACTTCAATGATGCTGCAAGAGGCTGATATGAAATGTCACATGTA 1850  
 Db 1520 TTTATGATGCTACAGCTGTGAGCAGCAGAGACTAAGCAATTTCAATATGAGACATGTA 1579  
 Qy 1851 TCAATTTGATCTTCTTACATGATGATGATGATGATGATGATGATGATGATGATGATG 1910  
 Db 1580 TCAATTTGATTTGCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 1639  
 Qy 1911 GTTGTGGAAATCTGCAAGAGATTTCTTTTGTGATGATGATGATGATGATGATGATGAT 1970  
 Db 1640 GTGTAGAAACCTGGGCTTGCACCTCATCTTTAA--TGAATAAATATGAAATATTA 1696  
 Qy 1971 CACAGCTCTAGTAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2030  
 Db 1697 CAAAGATTTGTGATCTTTCTTGTGAGAGCTTAAACAAAGAAATGCTTTTGGTGA 1756  
 Qy 2031 AAATTGCTTATGAT 2070  
 Db 1757 ATATGCTTATGAAACACACCTTAAAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1796  
 RESULT 7  
 AEX62951

ID ABX62951 standard; cDNA; 4423 BP.  
 XX ABX62951;  
 AC  
 XX  
 XX 25-FEB-2003 (first entry)  
 DT  
 XX  
 XX Human activated T cell cDNA #67.  
 DE  
 XX  
 XX T cell; gene; ss; differential expression; T cell activation;  
 KW antiallergic; cytosolic; immunosuppressive; antimicrobial; gene therapy;  
 KW allergy; cancer; graft versus host disease; infection;  
 KM autoimmune disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US2002137077-A1.  
 XX  
 XX 26-SEP-2002.  
 PD  
 XX  
 XX 25-OCT-2001; 2001US-00002600.  
 PF  
 XX  
 XX 25-OCT-2000; 2000US-0243521P.  
 PR  
 XX  
 XX (HOPK/) HOPKINS C M.  
 PA (PETE/) PETERSON D P.  
 PA (COCK/) COCKS B G.  
 PA (HAWK/) HAWKINS P R.  
 XX  
 XX Hopkins CM, Peterson DP, Cocks BG, Hawkins PR,  
 PI  
 XX WPI; 2003-102381/09.  
 XX  
 XX  
 PT New combination comprising several cDNAs that are differentially  
 PT expressed in activated T cells, useful for diagnosing, treating, staging  
 PT or monitoring treatment for allergy, cancer, infectious and/or autoimmune  
 PT disorders.  
 PT  
 XX  
 PS Claim 1; Page; 180bp; English.  
 PS  
 XX  
 CC This invention relates to the sequences of several cDNAs that are  
 CC differentially expressed in activated T cells. The sequences of the  
 CC invention may have antiallergic, cytosolic, immunosuppressive and  
 CC antimicrobial activity and may be used in gene therapy. The invention  
 CC also comprises a method for screening samples for differentially  
 CC expressed genes and a method for detecting these cDNAs by hybridisation.  
 CC The methods and compositions of the present invention are useful for  
 CC diagnosing, treating, staging or monitoring treatment for allergy,  
 CC cancer, chronic graft versus host disease, infectious and/or autoimmune  
 CC disorders. The present sequence represents a cDNA of the invention that  
 CC is differentially expressed in activated T cells  
 CC  
 XX  
 SQ Sequence 4423 BP; 1334 A; 740 C; 996 G; 1353 T; 0 U; 0 Other;  
 Query Match 17.3%; Score 384.8; DB 7; Length 4423;  
 Best Local Similarity 58.6%; Pred. No 4.3e-94;  
 Matches 762; Conservative 0; Mismatches 502; Indels 36; Gaps 4;

QY 1041 TCCTACCAATTTGGCTCA-----TATGATGATGATG 1073  
 DB 776 TTTTACCATCTAGTCAGATATATACAGATGCTCCAGAGAAAGCTTTGAAGCTGTGA 835  
 QY 1074 GAATPACTGCCAGTCGTTTAAAGATTGACAGAACCAAGTGTATTATTGTAGACACAA 1133  
 DB 836 AGAAATGGAAGGTATGGGCGCGCAACAAATATCCAAATTCCTTGTTTATGCCCCAA 895  
 QY 1134 CTCGAGATTTGCTAACCAAGTTTATTGGAAGCCAAATTTCTTTGGGACTGTG 1193  
 DB 896 CAAGAGATTTGCTGACAGATCTTAGAGAGCCAAATTTTCTTCCAGATCTGAG 955  
 QY 1194 TAAGAGCTGTTGTTATATATGAGGGAACCCAGCTGGACATTCATTTGACAAATAGTAC 1253  
 DB 956 TTGCTCCTTGTTAGTTTATGTTGTGTGTGATTTGTCAGACAGATTCGGGACTTAGAAC 1015  
 QY 1254 AAGCTGTATATATTTATGTCCTCTCTGGAAGACTGATGATATCATAGGCAAGAA 1313  
 DB 1016 GTGATGACCACTGTTGTATGACCACTCAGACGCTTAGATATATGAGAAAGAGAA 1075  
 QY 1314 AGATTGCTCAAAACAGATCAATTAATTAATTTGATGGAAGCTGATGCAATGTTGATA 1373  
 DB 1076 AGATTGATTAAGCTTCTGCAAGATCTTAGGTTGATGAGAGCTGATAGATGCTGATA 1135  
 QY 1374 TGGGTTTGTCTCAAGAAATGAAGATTATTTCTTGCCAGGAATGCCATCAAGAAAC 1433  
 DB 1136 TGGGATTTGAACCTCAGATACGTCGTATAGTTGAACAAGATTAATCTTACCAAAAGGCG 1195  
 QY 1434 AGCGCAACCCCTATATGTTGTCGCAACTTTCCAGAGGAATTCAAAGTTGGCTGAG 1493  
 DB 1196 TTGCTACACATGATGTTTATGCTACTTTTCTTAAGGAATACAGATGCTTGTCTGTG 1255  
 QY 1494 AGTTTAAAGTCAATTAATCTGTTTGTGCTGTGCAAGTGGTGGACATGTAGAG 1553  
 DB 1256 ACTTTTGGATGAA---TATATCTTTTGGCTGATAGGAGAGTGGCTCTACCTGTGAGA 1312  
 QY 1554 ATGTTACAGACACCGTTCTCCAAAGTGGCCAGTTCTCAAAAGGAAAGAAAGCTCGTTGAA 1613  
 DB 1313 ACATCACACAGAAAGTATTTGGTGGAAAGCTTAGATTAACGATCTTCTACTGACACA 1372  
 QY 1614 TTCTCGAAGCACTAGGG---GATGAAGAACTATGCTTGTGTAAGTAAGAAAAAG 1670  
 DB 1373 TTTTAGTGCAACAGAGAGTATTCATTAATTTAGTTGTGGAAGCCAAAAAGGAG 1432  
 QY 1671 CAGATTTTACGCACTTTTCTTGTCAAGAAAAATATCAACTACAGATTCATGTG 1730  
 DB 1433 CAGATTTCCCTGGAGATTTCTTATACATGAAGATATGCTGTATCTGATTTATGAG 1492  
 QY 1731 ATCGGGAACAGAGAGGAGGAGCAAGCTCTTGAGATTTTCGTTGAAAGTGCAG 1790  
 DB 1493 ACCGCTCACAGAGATGAGAGAGAGGCCCTTACACAGATTCGCTCAGAAAAAGCCAA 1552  
 QY 1791 TTCTTGTTGCTACTTCAGTACCTGACAGAGGCTGATATTTGAATGTCACATGTTA 1850  
 DB 1553 TTCTAGTGGTACAGCTGTGACAGACAGAGATAGACATTTCAATGATGACATGTTA 1612  
 QY 1851 TCAATTTGATTTCTTCTTACCAATTTGAATATGTTCACTGATTTGGGGTCTGTG 1910  
 DB 1613 TCAATTTGATTTGCAAGATATTTGAAGATATGTCATGATTTGGCCGTACAGAC 1672  
 QY 1911 GTTGGGAATCTGACAGAGCAATTTCTTTTGTATCTTGAATCGATTAACATTTAG 1970  
 DB 1673 GTGAGGAACCTGGGCTGTGCACTTATCTTTAA---TGAAAAAATATGAATATTA 1729  
 QY 1971 CACAGCTCTAGTAATAAGTATTTGACAGATGCTCAACAGAGTCTCTGATGTTGGAAG 2030  
 DB 1730 CAAGAGATTTGTGATCTCTTGTAGAGCTTAACAAAGAGTGTCTTGTGTTGAAA 1789  
 QY 2031 AATTTGCTTTAGTACATATATCTTGTGCTTCAGTGTAG 2070  
 DB 1790 AATGCTTTATGACACACATACAGAGGTGGAGTGTG 1829

1074 GAATACTGCCAGTCGTTTAAAGAGTTGCAGGAACCAAGAGTGTATTATTGTAGCAACCA 113

XX

XX





Db 1783 AATGGCTTATGACACCACTACAGGCTGCGTGG 1822

## RESULT 11

AB199243  
ID AB199243 standard; cDNA; 2187 BP.

AC AB199243;

DT 07-MAR-2002 (first entry)

XX Mouse ischaemic condition related cDNA sequence SEQ ID NO:67.

XX Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;

XX vasospastic ischaemia; ischaemic condition; ischaemic disease; ss.

OS Mus musculus.

PN M020018188-A2.

PD 22-NOV-2001.

PF 18-MAY-2001; 2001WO-JP004192.

PR 18-MAY-2000; 2000JP-00145977.

PA (UNINT.) UNIV NIHON SCHOOL JURIDICAL PERSON.

PI Ishikawa K, Asei S, Takahashi Y, Nagata T, Ishii Y;

DR WPI; 2002-034733/04.

PT P-PSDB; ABB57045.

XX Examining the ischemic condition (e.g. occlusive ischemia) by measuring

XX expression levels of particular genes defined in the specification or by

XX PT genes.

XX Claim 2; Page 210-215; 2690pp; English.

XX The present invention describes a method for examining ischaemic

XX conditions, comprising measuring the expression levels of particular

XX genes (I) in a test sample or determining the expression profile of a

XX gene group in the sample comprising genes selected from (I). The method

XX is useful for examining the ischaemic condition (e.g. compressive

XX ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring

XX expression levels of particular genes (AB199202 to AB199912, encoding the

XX protein sequences in ABB57020 to ABB57374) or by determining the

XX expression profile of a gene group comprising these genes. The expression

XX levels or expression profiles produced by these genes are used as an

XX indicator when screening for ischaemic condition-improving drugs or

XX therapeutics for ischaemic diseases. AB19913 and AB19914 represent PCR

XX primers for a mouse ischaemic condition related sequence, which are used

XX in the exemplification of the present invention.

XX Sequence 2187 BP; 635 A; 409 C; 596 G; 547 T; 0 U; 0 Other:

XX Query Match 16.7%; Score 371; DB 6; Length 2187;

XX Best Local Similarity 57.7%; Pred. No. 1.8e-90;

XX Matches 759; Conservative 0; Mismatches 520; Indels 36; Gaps 4;

QY 786 TCTTTGCACTTACACACAGGCAATACTTCGCAATACGACATCTTCTGGAG 845

Db 472 TCTTTTGGAGCAATCTGGATTAACCTTGGAATATGAGACATTCGATCGAG 531

QY 846 TGTCTGACATGATGACACACCAACCAATTCGACTTTGAGAGCTAATCTCTGAGA 905

Db 532 CAACAGGCAACACATGCTCTCCACACATGAAAGTTTCAGATGAGATGGAGAGAA 591

QY 906 CACGATTAACAACTTCTCTAAAGCTGTTTACTTAAGCTTACCTCCGTGGAGAAATACA 965

Db 592 TTATTATGGAGAAACATTGAGCTTACTCGTTATACCTCGCCAACTCCAGTGGAGAGATG 651

QY 966 GTATTCCTATCATCTTGCAGAGACGAGATTGATGCTGTGCTCAAAACAGGCTGGGA 1025

Db 652 CTATTCCTATCATCAAGAGAGAGAGACCTTGATGCTGTGCTCAAAACAGGCTGGGA 711

QY 1026 AGACTGGGCTTTTCTCCCAAAATTTGGCTCATATGATGATGAG----- 1074

Db 712 AAACGTGAGCATTTCTGTGGCCATCTTGAGTCAGATCTATGCTGATGGTCCAGAGAG 771

QY 1075 -----AATACTGCCAGTGGTTTAAAGATTSCAGAAACAGATGTA 1118

Db 772 CTGAGAGGCTATAGAGAGAAATGAAATATGAGCCGCTGCAACATATCATCTCTT 831

QY 1119 TTATTTAGACCAACTGAGAAATTTGTCACACGATTTATTTGGAAGCCAGAAATTTT 1178

Db 832 TGTACTGCGACCAACGAGAGAAATTTGCGAGATCTATGAGAGAGCCAGAAATTTCT 891

QY 1179 CTTTGGGACTTGTGTAAAGCTGTGTATATATGAGGGGAAACCAAGTGGGACATTCAA 1238

Db 892 CATACCGATCTAGAGTCCGTCCTGCGTTATATGAGTGGCTGAAATTTGCCAGAGA 951

QY 1239 TTCCGCAATATGTAACAAGCTGTATATATATATGCTACTCTGGAAGACTGATGATA 1298

Db 952 TTCGAGACTTAGAACGTGATGCCACTTGTATAGACCACTCCAGAGCTTATGATGA 1011

QY 1299 TCATAGGCAAAAGAAAGATTGCTCTCAACAGATCAATATTTAGTTTGGATGAGCTG 1358

Db 1012 TGATGAGAGAGAGGAGATCGGATTAGACTTCTGCAAAATACCTGCTGTATATAGAGCTG 1071

QY 1359 ATGCGATGTTGATATGAGGCTTTTGGTCCAGAAATGAAAGTTATTTCTTGCCAGAA 1418

Db 1072 ACCGATGTTGATATGAGGCTTTTGGTCCAGAAATGAAAGTTATTTCTTGCCAGAA 1131

QY 1419 TCCATCAAAAGAAACAGGCGCAAAACCTTATGTTAGTGCACCTTTCCAGAGAAATTC 1478

Db 1132 TGCCGCCAAAGAGTGTCCGCCACACTATGATGTTAGTGCACCTTTCCAGAGAAATTC 1191

QY 1479 AAAGTTGCTGCGAGATTTTAAAGTCAATATATGTTGCTGTGCTGAGACAGTGG 1538

Db 1192 AGATCTGCGCCGATATTTCTTAGATG--AGTACATATTTCTGCTGTAGAGAGATGG 1248

QY 1539 GTGAGCATGTAGATGTTAGACGACGCTTCTCCAAAGTTGCGCAATTCCTCAAAAGAG 1598

Db 1249 GGTCTACTTGAGAGAACTACACAAAGAGTTGCTGAGAGATATGATGTTGTA 1308

QY 1599 AAAAGCTGTTGAATTTCTGCGAAACATAG--GATGAAAGACTATGCTTTGTTG 1655

Db 1309 CATTTCTGCTTACCTTTTAATGCAACAGCAAGATTCCTGACTGATGTTGTTGG 1368

QY 1656 AAAGTAAAGAAAGAGAGATTTTACTGCAACTTTCTTTGTCAGAGAAATATCACTA 1715

Db 1369 AATCCAAAGAGAGAGAGATTTGCTGAGAGATTTCTTATACATGAGATATGCTGTA 1428

QY 1716 CAAGTATTCATGATGATGAGGAGACAGAGAGCGGAGCAAGCTTTGGAGATTTGCT 1775

Db 1429 CCAGTATCCATGAGACCGTTCTCAGAGAGATGAGGAGAGCCCTTACCAAGTCCCT 1488

QY 1776 TTGGAAGTGGCCAGATTTCTGTTGCTACTTCACTAGTACGCGAGAGGCTGATTTGAA 1835

Db 1489 CAGGAGAAAGCCCAATTTCTAGTGCCTACAGAGTACAGAGAGAGATGATATTTCAA 1548

QY 1836 ATGTGCAATGTTATCAATTTTGAATCTTCTTCAACATGATGATATGTTCTGAA 1895

Db 1549 ATGTGAGCATGTTATATATTTTGAATGCTGATGATATGATATGATGATGATGAT 1608

QY 1896 TTGGGCTATCTGCTGTTGAGGATATCTGAGAGCAATTTCTTTTGTGTTGAT 1955

Db 1609 TGGGCGTACAGGCGGTGTGGAGAACTTGTGCTTCCACCTCAATCTTTAA--TGAA 1665

QY 1956 CGGATTAACATTTAGACAGAGCTCTAGTAAAGTATGACAGATGCTCAACAGATGTC 2015

Db 1666 GGAATTAATATCAAAAGATTTTACTGATCTTCTTGTGAGAGAGAAACAGAGATGC 1725

QY 2016 CTGCAATGTTGAGAGAAATGCTTTAGTACATATCTTCTGCTTCACTGTTAG 2070



Db 1726 CTTCTTGTTAGAGAACATGCTTTTGAACACCACTACAAAGGCTACAGTCGTGG 1780

RESULT 12  
ACD13409  
ID ACD13409 standard; cDNA; 3158 BP.  
XX  
XX ACD13409;  
XX  
XX 13-AUG-2003 (first entry)  
XX  
DE Human DNA encoding a p53 modifier, SEQ ID 81.  
XX  
XX Human, ss; gene; p53 modifier; cytosolic; cancer; cytosolic;  
XX antitumorigenic; antiapoptotic; p53 pathway; breast cancer; colon cancer;  
XX kidney cancer; lung cancer; ovarian cancer; angiogenesis; cell cycle;  
XX apoptotic disorder; cell proliferation disorder.  
XX Homo sapiens.  
XX  
XX WO200299122-A1.  
XX  
XX 12-DEC-2002.  
XX  
XX 03-JUN-2002; 2002WO-US017382.  
XX  
XX 05-JUN-2001; 2001US-0296076P.  
XX 10-OCT-2001; 2001US-0328605P.  
XX 15-FEB-2002; 2002US-0357253P.  
XX  
XX (EXEL-) EXELIXIS INC.  
XX  
XX Friedman L, Plowman GD, Belvin M, Francis-Liang H, Li D, Funke RP,  
XX WPI; 2003-156859/15.  
XX P-PDB; ABO07236.  
XX  
XX Identifying modulators of the p53 pathway for use in treating apoptotic  
XX or cell proliferation disorders, comprises screening for agents that  
XX modulate activity of a human ortholog of genes that modify the p53  
XX pathway in Drosophila.  
XX  
XX Example 2; Page 363-364; 678pp; English.  
XX  
XX The invention relates to identifying (M1) a candidate p53 pathway  
XX modulating agent, by contacting an assay system comprising a purified HM  
XX polypeptide (human orthologue of genes that modify the p53 pathway in  
XX Drosophila) or nucleic acid with a test agent under conditions, where but  
XX for the presence of the test agent, the system provides a reference  
XX activity, and detecting a test agent-biased activity of the assay system.  
XX Also included are modulating (M2) a p53 pathway of a cell (comprising  
XX contacting a cell defective in p53 function with a candidate modulator  
XX that specifically binds to a HM polypeptide comprising an HM amino acid  
XX sequence, where p53 function is restored), modulating (M3) a p53 pathway  
XX in a mammalian cell (comprising contacting the cell with an agent that  
XX specifically binds an HM polypeptide or nucleic acid) and diagnosing (M4)  
XX from the patient; (b) contacting the sample with a probe for HM  
XX expression; (c) comparing the results with a control; and (d) determining  
XX whether the comparison indicates a likelihood disease). (M1) is useful  
XX for identifying modulators of the p53 pathway. A probe for HM expression  
XX is useful for diagnosing breast, colon, kidney, lung and ovarian cancer,  
XX in a patient, where the cancer has greater than 25 % expression level.  
XX Modulators identified by (M1) are useful in a variety of diagnostic and  
XX therapeutic applications, where disease or disorder prognosis is related  
XX to defects in the p53 pathway, such as, angiogenesis, apoptotic or cell  
XX proliferation disorders (e.g. cancer). Another two new methods (M2 and  
XX M3) are useful for modulating the p53 pathway of a cell, thus restoring  
XX the p53 function of the cell, so that the cell undergoes normal  
XX proliferation or progression through the cell cycle. (M2) and (M3) are  
XX also useful for treating defects in the p53 pathway such as angiogenesis,  
XX apoptotic or cell proliferation disorders. The present sequence is an HM

CC nucleic acid encoding a p53 pathway modifying protein  
XX  
XX Sequence 3158 BP; 906 A; 607 C; 814 G; 831 T; 0 U; 0 Other;  
SQ

Query Match 16.7%; Score 371; DB 7; Length 3158;  
Best Local Similarity 57.7%; Pred. No. 2.2e-90;  
Matches 759; Conservative 0; Mismatches 520; Indels 36; Gaps 4;

QY 786 TCTTGGACATTATCAGACAGGCACTTTCGCAATATCGACATCTTCTGTGGAG 845  
DB 758 TCTTTTCGAGAGCAACACCTGGATTTATTTTGAAATATCGATTCCTCACTTGGAG 817  
QY 846 TGTCTGACATGATGACCAACACGAAATCTGACTTTTGAAGAGCTAATCTGTGACA 905  
DB 818 CACAGGCAACAACATGCTCCACATATGAAAGTTTCAGTGATGTTGAGATGGAGAAA 877  
QY 906 CACGAAATACAAATGCTTAAAGCTGTATATCTAAGCTTATCTCTGCAAAAATATCA 965  
DB 878 TTATCATGGGAAATATGAGCTTACTGTATATCTGCTCCCACTCCAGTGCAGAAAGATG 937  
QY 966 GTATTCCTATCATCTTGCAGAGAGAGATTTGATGCTGTGCTCAACAGAGCTGTGGA 1025  
DB 938 CTATTCCTATTTATCAAGAGAAAGAGACTGTATGCTTGTGCCAAACAGGCTTGGAA 997  
QY 1026 AGACTGGGCTTTCTCTCAATTTTGGCTCATATGATGATGATG----- 1074  
DB 998 AAACCTGACGATTTCTGTTCCCATCTGTAGTGAATTTATTCAGATGCTCCAGCGAG 1057  
QY 1075 -----AATACCTGCAAGTGTTTTAAAGTTGACGAGACAGAGTGA 1118  
DB 1058 CTTGAGGCGCATGAAGAAATGAAAGATATGGGCGCCGCAACATATCCATCTCCT 1117  
QY 1119 TTATTTGAGACCACTGCAGAATTGCTCAACCAATTTATTTTGAAGCCAGAAATTTT 1178  
DB 1118 TGGTATTAGCACCAAGAGAGTGGACATGATGATCAAGAGAGAGAGAGAAATTTT 1177  
QY 1179 CTTTGGGACTTGTGTAAGAGCTGTGTTATATATATGAGGGAACCCAGCTGGACATTTCA 1238  
DB 1178 CATACGATTTAAGTTTCTGCTGCTGTGTTATATGTTGCTGCCGATATTTGTCACAGAA 1237  
QY 1239 TTGCACAATATGATCAAGGCTGTATATATATATATGCTCACTCTGGAAGACTGATGATA 1298  
DB 1238 TTGAGACTTGGAAAGCTGATGCAATTTGTATGATGCCACTCAGAGAGCTATGTGATA 1297  
QY 1299 TCATAGGCAAGAAAGATTTGCTCAACACATCAAAATATCTAGTTTGGATGAAAGCTG 1358  
DB 1298 TGAATGAAAGAGAAAGATTTGATGACTTTGCAAAATATCTGTGTAAGATGAGCTG 1357  
QY 1359 ATGCATGTTGATATGAGTTTGTGTCAGAAATGAAGATTAATTTCTGCCAGAA 1418  
DB 1358 ATCGATGTTGATATGAGGTTTGAAGCTTCAGATGTTGAAATATGCAACAAGATACTA 1417  
QY 1419 TGGCATCAAGAAAGACGCGCAACCTTATGTTCAAGTGAACCTTTCCAGAGAAATTC 1478  
DB 1418 TGCTTCAAAGGCTGCGCCACACATATATGTTAGTCTTCTTAAGGAAATAC 1477  
QY 1479 AAAGTTGCTGAGAGATTTTAAAGTCAAAATATCTGTTGTGCTGTGAGCAAGTGG 1538  
DB 1478 AAGATGTGCTGCTGATTTCTTAAATGA---TATATCTTGTGCTGTGAGAAAGTGG 1534  
QY 1539 GTGAGCATATGATGATGTTTCAGCAAGCCGTTTCCAAAGTTGCCAATTCAAAAGAG 1598  
DB 1535 GCTCTACCTCTGAAACATCAACAGAAAGTATGTTGGGTGGAAGAAATCCAGCAACGAT 1594  
QY 1599 AAAAGTGTGAAATTTGCGCAACATATAG---GAGTGAAGAAATATGCTGCTTGTGG 1655  
DB 1595 CATTTTGTGATCTCTTAATATGCAACGAGAGATTCAGTGAACCTTAGTGTGTTGG 1654  
QY 1656 AAACCTAAGAAAAAGAGATTTTACTGCAACTTTTCTTGTCCAGAAAAAATATCACTA 1715  
DB 1655 AAGCCAAAAGGTCAGATTTCTGAGAGATTTCTTATACATGAAGATATGAGATGTA 1714  
QY 1716 CAAGTATCATGCTATGCGGAACAGAGAGCGGAGCAACGCTTGTGAATTTTCCT 1775

DB 1715 CCAGCATCCAGAGACCCCTTCTCAGAGGATATAGAGAGAGGCCCTTCAACCACTTCCGCT 1774  
 QY 1776 TTGAAAGTGGCCACTTCTTGTGACTTACAGTACGTCAGAGGCGCTGATATTGAA 1835  
 DB 1775 CAGGAAAAAGCCCAATTTAGTGTACAGCAGTAGCAGCAAGGAGCTGACATTTTCA 1834  
 QY 1836 ATGTCAACATGTATTCATATTTTGTCTTCTTACCATGATGATATGTCATGAA 1895  
 DB 1835 ATGTAAACATTTTCAATTTTGTGCTTCCCAAGTATTTGAAGATATGTACATGTA 1894  
 QY 1896 TTGGGCGTACTGTCTGTGTGGAAATACGCGACAGCAATTCCTTTTGTATCTTAAT 1955  
 DB 1895 TTGGTGTACGCGAGGTAGAGAACTTGGCCCTGGCAACCTCAATCTTTAA--CGAGA 1951  
 QY 1956 CGGATACCATTTAGCAGAGCCTCTAGTAAAGATTGACAGATGCTCAAGAGATTC 2015  
 DB 1952 GGAACATTAATTTACTTAAGATTTGTGTGATCTTCTTGTGAGCTAAACAGAAATGC 2011  
 QY 2016 CTGCATGTTGGAAGAAATTCCTTTAGTACATACATTCCTGCTTCACTGTAG 2070  
 DB 2012 CCTCTGTGTGAAAAACATGCTTATGACACCACTCAAGGCTAGCAGTCTGG 2066

RESULT 13  
 AA69632  
 ID AA69632 standard; DNA, 3408 BP.  
 AC AA69632;  
 XX  
 DT 02-FEB-1999 (first entry)  
 DE Dead Box X (DBX) gene short transcript nucleic acid sequence.  
 XX  
 KW Non-recombining region; human; Y chromosome; X homologue; testis; DBX;  
 KM infertility; sperm; gene alteration; inhibitor; Dead Box X; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO9846747-A2.  
 XX  
 PD 22-OCT-1998.  
 XX  
 PF 10-APR-1998; 98WO-US007115.  
 XX  
 PR 11-APR-1997; 97US-0041877P.  
 XX  
 PA (WHED) WHITEHEAD INST BIOMEDICAL RES.  
 XX  
 PI Lam BT, Page DC;  
 XX  
 DR WPI; 1998-568729/48.  
 DR P-PSDB; AAW81502.  
 XX  
 PT Novel genes in the non-recombining region of Y chromosome - useful to  
 PT diagnose if male infertility or reduced sperm count has a genetic basis.  
 PS  
 PS Disclosure; Fig 3A-B; Sapp; English.  
 XX  
 CC This represents the nucleotide sequence of the Dead Box X (DBX) gene  
 CC short transcript. The invention relates to genes occurring on the non-  
 CC recombining region of the human Y chromosome. The sequences fall into two  
 CC classes: (1) X-homologous DNA which are expressed in many organs, having  
 CC functional X homologues and (2) testis-specific DNA sequences. Y  
 CC chromosomal DNA from males with known conditions such as infertility and  
 CC reduced sperm count can be assessed using the invention to determine  
 CC whether the condition is associated with or caused by the occurrence of  
 CC the gene or gene alteration. Candidate inhibitors of the enzymatic  
 CC activity of the genes can be assessed using in vitro assays  
 XX  
 SQ Sequence 3408 BP; 951 A; 734 C; 923 G; 800 T; 0 U; 0 Other;  
 Query Match 16.6%; Score 369.4; DB 2; Length 3408;

Best Local Similarity 57.6%; Pred. No. 6.1e-90;  
 Matches 758; Conservative 0; Mismatches 521; Indels 36; Gaps 4;  
 QY 786 TCTTTGCACTTATCAGACAGCATTAATCTTCGAAATATGACACTATTTCTTGGAG 845  
 DB 1305 TCTTTTCTGAGGCAACCTGGATTAATTTTGAATATCGATGACATTTCCAGTTAG 1364  
 QY 846 TGTCTGACATGATGACACAGCAATTCGACTTTTGAAGAGCTAATCTCTGTCA 905  
 DB 1365 CAACAGGCAACACTGTCTTCCACATATTTGAAGTTTCAATGATTTGAATGGAGAA 1424  
 QY 906 CACTGATTAACAATTTGCTTAAGCTGTATTAATTAAGTTTCTCTGTGCAAAATACA 965  
 DB 1425 TTATCTGAGGAAATATGAGCTTATCTGTATCTCCCACTCCAGTCCAAAGCATG 1484  
 QY 966 GATTCCTATCATCTTGCAGAGACGATTTGATGCTGTGCTCAACAGGCTTGGGA 1025  
 DB 1485 CTATTCCTTATTAAGAAAGAAAGACATTTGATGCTTGTGCCAAACAGGCTTGGAA 1544  
 QY 1026 AGACTGCGCTTTTCTCCATCAATTTGGCTATATGATCATATG-- 1074  
 DB 1545 AAACGACAGCAATTTCTGTGCTCCATCTTGAATGATTTTGAAGTGTCCAGGAG 1604  
 QY 1075 -----AATACTGCGCTGTTTAAAGTTTCAGAGAACAGATGTA 1118  
 DB 1605 CTTGAGGCGCATGAGAAATGAAAGTATGGGCGCCCAACATACCAATCTCT 1664  
 QY 1119 TTATTTGACCACTCGAGAAATTTGTCACACGATTTATTGGAAGCCAGAAATTTT 1178  
 DB 1665 TGTATTGACCACTCGAGAGAGTTGGAGTACAGATCTAGCAAGAACGAGAAATTTT 1724  
 QY 1179 CTTTGGGACTTGTGTAAGCTGTGTATATATGGGAGAACCCGCTGGGCAATTC 1238  
 DB 1725 CATACCATTTAGAGTCTCTTGTGCTGTATATGATGATGATGATGATGATGATG 1784  
 QY 1239 TTGCAAAATAGTACAAAGCTGTATATATATATGATGATGATGATGATGATGATG 1298  
 DB 1785 TTGAGACTTGGAGAGTGGATGCTATTTGATAGGCACTCCAGAGCTCTGTGAT 1844  
 QY 1299 TCATGAGGAAAGAAAGATTTGCTCAACAGATCAATTAATTTGATGAGAGCTG 1358  
 DB 1845 TGATGAAAGAGGAAAGATTTGATTAATTTGCAATTTCTTGATGATGATGATGATG 1904  
 QY 1359 ATTCGATTTGATATGATGATTTTGTGCTCAGAAATGAAGATTAATTTCTGCGCAG 1418  
 DB 1905 ATTCGATTTGATATGATGATTTGATGCTCAGATTCGTAATATGTAAGATTAAT 1964  
 QY 1419 TGCCATCAAGAGACAGGCGCAACCTTATGCTCAGTGAATTTTCCAGAGAAATTC 1478  
 DB 1965 TGCTCCAAAGGAGTGTGCTGCGCAGCTATGATTTTATGCTATCTTCTTCAAGAAAT 2024  
 QY 1479 AAAGTTGCTGCAAGATTTTAAAGTCAATTAATCTGTTGTGCTGTGCAAGATG 1538  
 DB 2025 AGATGCTGCTGATGATTTCTTAATGA---TATATCTTGTGCTGTGAGAGGTTG 2081  
 QY 1539 GTGAGCATATGATGATTTTCAAGACCTTCTCAAGTGTGCGAGTTCTCAAAAGAG 1598  
 DB 2082 GCTTACTCTGAAATATCAGACAGAAATGATTTGGTGGAGATTCAGAAACGCT 2141  
 QY 1599 AAAGCTGTTGAATTTCTGGAATATAG--GATGAAGAACTATGCTTTGTTG 1655  
 DB 2142 CATTTCTGCTGACCTCTTAATATCAACAGCAAGATATCACTGACCTTGTGTTGTG 2201  
 QY 1656 AAACGTAAGAAAGAGATTTTACTGCACTTTTCTTCAAGAAATATCAACTA 1715  
 DB 2202 AGACCAAAAGAGGAGATTTCTTGAAGATTTCTTAACATGAAGATTAAGCATGTA 2261  
 QY 1716 CAAGTATCATGATGATGAGAGAGAGAGGAGAGAGCTTGTGAGATTTGCT 1775  
 DB 2262 CCAGCATCATGAGAGAGCTTCTCAGAGGATGAGAGAGGCTTCAACCAATTCGCT 2321  
 QY 1776 TTGAAATGCGGCTTCTTGTGCTATCTTCAATGCTGCAAGGCTGATATGAA 1835

Db 2322 CAGAAAAAGCCCAATTAGTGTCTACAGCAGTAGCAGACAGAGACTGCATTTTCAA 2381  
 Qy 1836 ATGTCACACAGCTTATCAATTTTATCTTCTCTTACCATTTGATGATGTTTCATGAA 1895  
 Db 2382 ATGTAAACATGATTAATCAATTTTATCTTCTCTTACCATTTGATGATGTTTCATGAA 2441  
 Qy 1896 TTGGCCTGATCTGCTGTGTGGGAATATCTGACAGACAAATTTCTTTTGTGATTTGAAT 1955  
 Db 2442 TTGGCTGATGAGGAGCTGTAGGAAACCTTGGCTGGCAACTCTATTCTTAA--CGAGA 2498  
 Qy 1956 CGGATTAACCATTTAGCAGACGCTCTAGTAAAGTATTGACAGATGCTCAACAGGATGTC 2015  
 Db 2499 GGAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2558  
 Qy 2016 CTGCATGCTGTAAGAAATGCTCTTATGATACATTCCTGCTTACGTTAGTGTA 2070  
 Db 2559 CGCTGTGTTAGAAACATGCTTATGACACCCCTCAAGGGTAGCAGTGTGG 2613

RESULT 14  
 ID AAA29207 standard; DNA; 5321 BP.  
 AC AAA29207;  
 XX  
 DT 12-SEP-2000 (first entry)  
 DE Human DBX1 coding sequence.  
 XX  
 KM Topoisomerase III- $\alpha$ ; interacting protein; TIII- $\alpha$  IP; RNA helicase;  
 KM DBX1; nucleosome; supercoiling; chromosome segregation; recombination;  
 KM stabilisation; cell division; apoptosis; cell cycle regulation;  
 KM cytotoxic; anti-tumour; DBX1; ds.  
 XX Homo sapiens.  
 OS  
 FH Key Location/Qualifiers  
 FT CDS 856..2844  
 FT /tag= a  
 FT /standard\_name= "DBX1"  
 FT /product= "RNA\_helicase-like\_protein"  
 PN MO200032768-A1.  
 PD 08-JUN-2000.  
 XX  
 PF 29-NOV-1999; 99WO-FR002952.  
 XX  
 PR 30-NOV-1998; 98FR-00015081.  
 XX  
 PA (AVENTIS PHARMA SA.  
 PI Fournier A, Goulaouic H, Riou J;  
 XX WPI, 2000-412316/35.  
 DR P-PSDB; AAY96483.  
 XX  
 PT New nucleic acid encoding ligand for topoisomerase III  $\alpha$ , useful for  
 PT inhibiting the enzyme and in drug screening, e.g. for potential  
 PT anticancer agents.  
 XX  
 PS Claim 6; Page 47-53; 68pp; French.  
 CC This DNA encodes a DBX1, which shows homology with RNA helicases but the  
 CC activity of a helicase has never been demonstrated and its function has  
 CC not yet been identified. DBX1 possesses the 8 characteristic motifs of  
 CC the helicases of the family "DBX". In particular, it appears to be part  
 CC of the sub-family represented by helicase P110. The DBX1 gene is situated  
 CC on the X chromosome and its homologue, which is situated on the Y  
 CC chromosome has 91 percent identity with a novel protein coding sequence.  
 CC The novel protein (see AAY96482) is a human topoisomerase III- $\alpha$  IP  
 CC interacting protein (TIII- $\alpha$  IP). TIII- $\alpha$  IP has structural features in  
 CC common with RNA helicases (e.g. DBX1), which are involved in destruction

CC of the nucleosome structure, supercoiling of DNA, segregation of newly  
 CC replicated chromosomes, and recombination and stabilisation of the  
 CC genome. Agents that alter interaction between TIII- $\alpha$  IP and TIII- $\alpha$   
 CC therefore modulate cell division, replication, transcription,  
 CC translation, splicing and DNA recombination or repair, so may slow cell  
 CC growth, block the cell cycle or induce apoptosis. Antibodies and ligands  
 CC of TIII- $\alpha$  IP are used to prevent, treat or alleviate diseases that  
 CC involve abnormal regulation of the cell cycle, i.e. they are potential  
 CC anti-tumour (cytotoxic) agents  
 XX

Sequence 5321 BP; 1532 A; 1010 C; 1274 G; 1505 T; 0 U; 0 Other;  
 Query Match 16.6%; Score 369.4; DB 3; Length 5321;  
 Best Local Similarity 57.6%; Pred. No. 7.6e-90;  
 Matches 758; Conservative 0; Mismatches 521; Indels 36; Gaps 4;

Qy 786 TCTTTCACATTAATGACAGCAGCAATTAATCTGCAAAATAGACACTATTTCTTGAGAG 845  
 Db 1304 TCTTTCGAGGCAACCTGAGATTAATTTGCAAAATAGACACTATTTCAATTTAGG 1363  
 Qy 846 TGTCTGACATGATGACACACACCAATTCGACTTTGAAAGAACTAATCTGTGAGA 905  
 Db 1364 CAACAGGCAACACCTGCTCCACATATTAAGTTCAAGTATGATGATGGAGAAA 1423  
 Qy 906 CACTGAATTAACATGCTTAAAGCTGTTAATTAAGTTAATCTCTGTGCAAAATATCA 965  
 Db 1424 TTAATGAGGAAACCTTGAAGCTTAATCTGCTTACCTCCCACTCCAGTCAAAAGCATG 1483  
 Qy 966 GTATTCCTATCATCTTGACAGCAGCAATTTGATGCTGTGCTCAACAGAGGCTGAGA 1025  
 Db 1484 CTATTCCTATTAATCAAGAAAGAAAGAACTGATGCTGTGCTCAACAGAGGCTGAGA 1543  
 Qy 1026 AGACTGCGGCTTTCTCTCTACCAATTTTGCTCATATGATGATGATG----- 1074  
 Db 1544 AAACGTGACGATTTCTGTCCCACTTGAGTATGATGATGATGATGATGATGATGATGATG 1603  
 Qy 1075 -----AATACTGCGAGCTGTTTAAAGTGTGCGAAGACGAGTGA 1118  
 Db 1604 CTTGAGGCGCATGAAGAAATATGAAAGATATGAGGCGCCCAACATATCCATATCTCT 1663  
 Qy 1119 TTATTTAGACACCACTGCGAATTTGCTCAACAGATTTATTTGGAAGCCAGAAATTTT 1178  
 Db 1664 TGGTATTAAGACCAACAGAGAGCTGCACTATGATGATGATGATGATGATGATGATGATG 1723  
 Qy 1179 CTTTGGGACTGTGTAGAGCTGTGTATATATATGAGGAGAACCCAGCTGGGCAATTCGA 1238  
 Db 1724 CATACCATTAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1783  
 Qy 1239 TTGACAAATATGACAGGCTGTAATATATATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1298  
 Db 1784 TTGAGACTTGGAACTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1843  
 Qy 1299 TCATAGGCAAGAAAGATTTGCTCAACAGATTCATTAATTTAGTTTGGATGAAGCTG 1358  
 Db 1844 TGAATGAAGAGAAAGATTTGATGCTTGTGCAAAATCTGTTGTTGATGATGATGATGATG 1903  
 Qy 1359 ATGCAATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1418  
 Db 1904 ATGCAATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1963  
 Qy 1419 TGGCATCAAGAAAGAGCCCAACCTTATGTTAGTGAACATTTTCCAGAGAAATTC 1478  
 Db 1964 TGGCTCAAGAGGTGTCCCACTATGATGATGATGATGATGATGATGATGATGATGATGATG 2023  
 Qy 1479 AAAGTTGGCTCAGAGTTTAAATCAATTTCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTG 1538  
 Db 2024 AGATGCTGCTGCTGATTTCTTATGATGA---TATATCTTCTTGGCTGATGAAAGTTG 2080  
 Qy 1539 GTGAGCATGTAGAGATGTTGACGACCTGTTCCAAAGTTGGCCAGTTTCCAAAGAG 1598  
 Db 2081 GCTTACCTCTGAAACATCAACAGAAAGATGTTGGTGAAGATCAACAAACGCT 2140  
 Qy 1599 AAAGCTGCTGAAATTTCTGGAACATGAG---GATGAAGAACTATGCTTTTGTG 1655

Db	2141	CAATTCGCTTGACCTCCTAAATGCACACAGGCAAGATTCAGTGCACCTTGATGTTGTGG	2200
Qy	1656	AAACTAAGAAAAAAGCAGATTTTACTGCAACTTTCTTTGTCAAGAAAAATATCAACTA	1715
Db	2201	AGACCAAAAAAGGGTGCAGATTTCTCTGAGGATTTCTATACATGAAGATACGCATGTA	2260
Qy	1716	CAACTATCCATGTGTGATCGGGAGACAGAGAGCGCGGAGCAAGCTCTTGAGATTTTCGCT	1775
Db	2261	CCAGCATTCATGAGAGACCGTTCTCAGAGGGATAGAGAAAGGCCCTTCACCAAGTCCGCT	2320
Qy	1776	TTGGAAGTGCCTCCAGTCTTGTTGTCATCTTCAGTAGCTGCCAGAGGCGCTGATTTGAA	1835
Db	2321	CAGGAAAAAGCCCAATTTTACTGTGCTACAGCAGTAGCAGCAAGAGAGACTGGACATTTCAA	2380
Qy	1836	ATGTGCACATGTATATCAATTTTGTATCTTCCTTACCATGATGAAATATGTCATGGA	1895
Db	2381	ATGTGAACAACATGTTATCAATTTTGACTTGCCCAAGATATTTGAAGATATGTACATGTA	2440
Qy	1896	TTGGGCGCTACTGTCGCTGTGTCGGAATACCTGCGCAGAGCAATTTCCTTTGTATCTGAAT	1955
Db	2441	TTGGTTCGACGGGAGCGTGTAGGAACCTTGCGCTGGCAACTCAATCTTTAA--CGAGA	2497
Qy	1956	CGGATTAACCATTTTAGCACACGCTCTAGTAAAGATTACACAGATGCTTAACAGATGTC	2015
Db	2498	GGAACATATAATTACTAAGAGATTGTGTGATCTTCTGTGTAAGCTAAACAGAAAGTGC	2557
Qy	2016	CTGCATGTTGTGAAGAAATTTGGCTTTAGTACATACATATCTCGCTTCAGTGGTAG	2070
Db	2558	CGCTTGTTAGAAACATGCTTATGAAACACACACTAACAAGGTTAGCAGTGTGTG	2612

RESULT 15  
AAV69631  
ID AAV69631 standard; DNA; 5322 BP.

XX	AAV69631;	
AC		
XX	02-FEB-1999 (first entry)	
DT		
XX		
DE	Dead Box X (DBX) gene long transcript nucleic acid sequence.	
XX		
KM	Non-recombining region; human; Y chromosome; X homologue; testis; DBX;	
XX	infertility; sperm; gene alteration; inhibitor; Dead Box X; ss.	
KW		
XX		
OS	Homo sapiens.	
XX		
PN	W09846747-A2.	
XX		
PD	22-OCT-1998.	
XX		
PF	10-APR-1998; 98WO-US007115.	
XX		
PR	11-APR-1997; 97US-0041877P.	
XX		
PA	(WHED) WHITEHEAD INST BIOMEDICAL RES.	
XX		
PI	Lahn BT, Page DC;	
XX		
DR	WPI; 1998-568729/48.	
XX		
DR	P-PSDB; AAW81501.	
XX		
PT	Novel genes in the non-combining region of Y chromosome - useful to	
XX	diagnose if male infertility or reduced sperm count has a genetic basis.	
PS	Disclosure; Fig 3A-B; 54pp; English.	
XX		

This represents the nucleotide sequence of the Dead Box X (DBX) gene long transcript. The invention relates to genes occurring on the non-recombining region of the human Y chromosome. The sequences fall into two classes: (1) X-homologous DNA which are expressed in many organs, having functional X homologues and (2) testis-specific DNA sequences. Y chromosomal DNA from males with known conditions such as infertility and

CC reduced rpm count can be assessed using the invention to determine  
CC whether the condition is associated with or caused by the occurrence  
CC of the gene or gene alteration. Candidate inhibitors of the enzymatic  
CC activity of the genes can be assessed using in vitro assays  
xx  
SQ Sequence 5332 BP; 1532 A; 1011 C; 1274 G; 1505 T; 0 U; 0 Other;

Query Match	16.6%;	Score 369.4;	DB 2;	Length 5322;
Best Local Similarity	57.6%;	Pred. No. 7.6e-90;		
Matches 758;	Conservative 0;	Mismatches 521;	Indels 36;	Gaps 4;

786 TCITTCGACATTATCGAGACGGCATAAAGCTTCGACAAATACGACACTATTCTTGCGANG 845

Db 1305 TCITTTCTGAGGCAACACTGGATTATTTTGAGAATACGATGACATTCCAGTTGAGG 138

846 TGTCTGGACATGATGCACCCAGCAGCAATTCGTGCTTTGAAGAGAGCAATCCTGTCAGG 503

[illegible][illegible]

DB 1473 11A1C816GGAACCA110NOC.....CC.....CG.....

27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841  
842  
843  
844  
845  
846  
847  
848  
849  
850  
851  
852  
853  
854  
855  
856  
857  
858  
859  
86

1026 AGACTGCGGCTTTCTCCTACCAATTTGGCTCATATGATGCATGATGC----- 107

1545 AACTGCAGATTCTGTGCCCATCTTGAGTCAGATTATTCAGATGATCCAGCGAG 160

1075 -----ATACTGCCAGTCGTTTAAAGAGTTGCAGGAACGAGAGTGA 111

Db 1605 CTTTGAGGGCCATGAAGAAATGAGTATGGGGCCCGCAACCAATACCCCAATCTCT 166

1119 TTATTGTAGCACCACCTCGAATTCGTCAACCAGATTATTGGAAGCCAGAAATTTT 117

Db  
1665 TGGTATTAGCACCAAGAGAGTTCGACGTACAGATCTACGAAGAGCCAGAAAATTTT 172

1179 CTTTGGGCACTGTCTAGAGCTGTGTTATATATGCGGGAACCCAGCTGGGACATTCA 123

Db 1725 CATACGATCTAGAGTTCGTCTTGCGTGTATTATGTTGGTGCCGATATTGGTACAGCAGA 178

1239 TTGGCAATAGTACAGGCTGTAAATATATTATGTCTACTCCTGGAGACTGATGGATA 129

Db 1785 TTGAGACTTGGACGTGGATGCCATTGTTAGTAGCCACTCCAGAGCTCTAGTGGATA 184

1299 TCATAGCAAGAAGATTGGTCTCAACAGATCAATACCTAGGTTTGGAGTGGGCTG 1300

D6 1845 TGATGGAGGAGGAAAGATGGATACAGCTTTTGCATTTCCTGTGGTGATGGAATGCAATCAGCAGAACCC 141

[illegible]

147  
1305 AAGGAGTATCCAAATACCACTTTTCCTTTTCCAGAAGGAATTTC

1965 TGGCTCCAGAGCGGTATCCGACACATATGATGTTAGTCACTTTTCCTAGGAAATAC 202

1479 AAAGGTTGGCTGCAGAGTTTTAAAGTCAAA TTATCTGTTTTGTTGCTGTGGACAAGTGG 153

Db 2025 AGATGCTGCTGTAATTTCTAGATGA---TATATCTTCTTGCGCTGAGAGAGTTG 209

1539 GTGAGCATGTAGAGATGTTACGACGACCGTTCTCCAAAGTTGGCCAGTTCTCAAAAAGAG 1550

Db 2082 GCTCTACCTTGAAAACATCACACAGAAAGTGTGGGTGGAGAATCAGACAAACGGT 214

1599 AAAGCTCGTTGAATTCCTCGAACAATAG---GGATGAAGAAGACTATGGTCTTTGTTG 165

Db 2142 CATTCTGCTGACCTCCCTAAATGCACAGCAAGATTCACTGACCTTAGTGTGTCG 220

1656 AAACTAAGAAAAAGCAGATTCTGCACTTTCTTTGTCAAGAAAAAATCACTA 17

Db 2202 AGACCAAAAAGGTCAGATTCCTCGAGAGATTTCTTATACATGAAGATACCATGTA 2261  
QY 1716 CAAGTATTCATGCTGATCGGGAACAGAGAGCGGAGCAAGCTCTTGAGATTTGCT 1775  
Db 2262 CCAGCATCCATGAGACCGTTCTCAGAGGATAGAGAAAGGCCCTTCACCACTCCGCT 2321  
QY 1776 TTGAAAGTCCCGAGTCTTGCTGCTACTTCAGTAGCTGCCAGAGGGCTGGATATGAA 1835  
Db 2322 CAGGAAAAAGCCCAATTTTATGCTACAGCAGAGCAGAGGACCTGACATTTCAA 2381  
QY 1836 ATGTGCAACATGTTATCAATTTTGATCTTCTTACCATTTGATATGTCATCGAA 1895  
Db 2382 ATGTGAAACATGTTATCAATTTTGACTTCCCAAGTGATTTGAAGATATGTACATGTA 2441  
QY 1896 TTGGCGTACTGTCGTTGCGGAATACCTGGCAGAGCAATTCCTTTTGTGATCTGAAT 1955  
Db 2442 TTGGTCGTACGGACGCTGTAGGAAACCTTGCCCTGGCACTCATTTCTTAA---CGAGA 2498  
QY 1956 CGGATTAACATTAGACACAGCCTCTAGTAAAGTATGACAGATGCTCAACAGATGTC 2015  
Db 2499 GGAACATTAATATTACTTAAGATTTGTGATCTTCTGTTGAAGCTAAACAAGAGTGC 2558  
QY 2016 CTGCATGCTGGAAGAAATGCTTATAGTACATACATTCCTGCTTCACTGCTAG 2070  
Db 2559 CGTCTTGTTAGAAAACATGGCTTATGAACACCACTAACAGGCTACAGTCTGG 2613

Search completed: July 27, 2004, 06:51:00  
Job time : 901 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: July 27, 2004, 03:01:17 ; Search time 8597 Seconds

(without alignments)  
11212.622 Million cell updates/sec

Title: US-09-714-865B-1

Perfect score: 2224  
Sequence: 1 actgaagtcacccatggggg.....aagctctgggtcttgatgca 2224

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6340544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
GenEmbl:
1: gb_da:
2: gb_hlg:
3: gb_in:
4: gb_cm:
5: gb_ov:
6: gb_pat:
7: gb_ph:
8: gb_pi:
9: gb_pr:
10: gb_ro:
11: gb_scs:
12: gb_sy:
13: gb_un:
14: gb_vi:
15: gb_da:
16: em_fun:
17: em_hum:
18: em_in:
19: em_mu:
20: em_om:
21: em_or:
22: em_ov:
23: em_pat:
24: em_ph:
25: em_pi:
26: em_ro:
27: em_scs:
28: em_un:
29: em_vi:
30: em_hlg_hum:
31: em_hlg_inh:
32: em_hlg_other:
33: em_hlg_mus:
34: em_hlg_pin:
35: em_hlg_rtd:
36: em_hlg_mam:
37: em_hlg_vrt:
38: em_sy:
39: em_hgo_hum:
40: em_hgo_mus:
41: em_hgo_other:
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2224	100.0	2224	AY004154	AY004154 Homo sapi
2	2214.4	99.6	2411	AF262962	AF262962 Homo sapi
3	1997.2	89.8	2363	BC047455	BC047455 Homo sapi
4	1939.6	87.2	2189	HSM802178	AL137462 Homo sapi
5	1601	72.0	1984	AX748127	AX748127 Sequence
6	1601	72.0	1984	AK093439	AK093439 Homo sapi
7	1594	71.7	3030	575275	S73275 RVG-vasa-1
8	1443	64.9	1930	MUSDVH	D14859 Mus musculu
9	751.2	33.8	1989	AB004836	AB004836 Gallus ga
10	712	32.0	2502	AF046043	AF046043 Xenopus l
11	686.8	30.9	2865	AB005147	AB005147 Danio rer
12	686.4	30.9	2492	DRY12007	Y12007 D. rerio vlg
13	651.4	29.3	2337	AF479820	AF479820 Cyprinus
14	619.4	27.9	2038	AB051835	AB051835 Oreochrom
15	619.4	27.9	2130	AB032467	AB032467 Oreochrom
16	612.4	27.5	2052	AB038252	AB038252 Leucopisc
17	612.2	27.5	2612	AB032566	AB032566 Oncorhnc
18	565	25.4	1512	AF479823	AF479823 Pantodon
19	552.8	24.9	2187	AB063484	AB063484 Oryzias l
20	548.2	24.6	1198	AF251800	AF251800 Danio dan
21	543.8	24.5	1730	AF479821	AF479821 Hyphessob
22	503.6	22.6	1586	AF520608	AF520608 Sparus au
23	503.4	22.6	1506	AF479824	AF479824 Melanocae
24	498.6	22.4	1816	AF479825	AF479825 Oncorhnc
25	493	22.2	504	BD110839	BD110839 EST and e
26	493	22.2	504	BD110839	BD110839 Sequence
27	487.4	21.9	2551	AB016603	AB016603 Clona int
28	487.4	21.9	2551	AB016603	AB016603 Clona int
29	457	20.5	1873	AB047385	AB047385 Ephydratia
30	456.6	20.5	1332	AB047383	AB047383 Hydra mag
31	438.6	19.7	1764	AF510054	AF510054 Schistoc
32	438	19.7	3023	AB047803	AB047803 Clona sav
33	436.4	19.6	2777	AB047802	AB047802 Clona sav
34	434.6	19.5	2721	AB047382	AB047382 Hydras mag
35	428.8	19.3	1449	AF479822	AF479822 Oryzias l
36	384.8	17.3	2713	BC034942	BC034942 Homo sapi
37	383.2	17.2	2319	HSARF000985	AF000985 Homo sapi
38	383.2	17.2	4416	HSARF000984	AF000984 Homo sapi
39	378.2	17.0	2384	AB047381	AB047381 Hydra mag
40	377	17.0	524	AY100475	AY100475 Equus cab
41	371	16.7	2187	AX305316	AX305316 Sequence
42	371	16.7	2187	10 MUSRNABELI	L25126 Mus musculu
43	371	16.7	2201	AF061337	AF061337 Homo sapi
44	371	16.7	2499	BC011819	BC011819 Homo sapi
45	371	16.7	3158	HSU50553	US0553 Homo sapien

## ALIGNMENTS

RESULT 1  
AY004154  
LOCUS AY004154 2224 bp mRNA linear PRI 25-AUG-2000  
DEFINITION Homo sapiens DEAD box RNA helicase (Vasa) mRNA, complete cds.  
ACCESSION AY004154  
VERSION AY004154.1 GI:9438226  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 2224)  
Castrillon,D.H., Quade,B.J., Wang,T.Y., Quigley,C. and Crum,C.P.  
The human Vasa gene is specifically expressed in the germ cell  
lineage

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9585-9590 (2000)  
 MEDLINE 20402578  
 PUBMED 10920202  
 REFERENCE 2 (bases 1 to 2224)  
 AUTHORS Castriello, D.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (04-JUL-2000) Pathology, Brigham and Women's Hospital, 75 Francis Street, Boston, MA 02115, USA  
 FEATURES  
 source Location/Qualifiers  
 1..2224  
 /organism="Homo sapiens"  
 /mol\_type="rRNA"  
 /db\_xref="taxon:9606"  
 /chromosome="5"  
 /map="5q11.2-12"  
 /tissue\_type="testis"  
 /dev\_stage="adult"  
 1..2224  
 /gene="VASA"  
 14..2188  
 /gene="VASA"  
 /codon\_start=1  
 /product="dead box RNA helicase"  
 /protein\_id="AA06585.1"  
 /db\_xref="GI:9438227"  
 /translation="MGDEWBAEINPMSSVVDIPEKDRYSGENEDNFRTPASSSEM  
 DDGSRSDHEPMKSGFASGRNFGNEDAGECNKRDNTSMGFGVKGSPGRGNSRFE  
 DSSGFWRESNDCEINPFRNRFSGRGYRDGNSSEAGPYRGRSGFPGRCGF  
 GLGSPNDLDDECMORTGIFGSRPVLSTGNGDTPSGSGSERGGYGLNIEV  
 ITGSKNSWKEAEAGSGSDTQPKXTVTPPPEDSDI PAHYQGNIDPDYDTIV  
 EYSGHDAFALITFEENICOTLANNIARKGYTKLTPVQKSIPIILARDIMACQT  
 GSGTAFLFLILAHMDGILARFELDEPCTIICATPGRLMDIIGEKIQLQIYLVLEDR  
 CVRAVYIGGTOLEHSIRQIVQCNIIICATPGRLMDIIGEKIQLQIYLVLEDR  
 MLDWGPEMKKLLISCPGMSKEQROTLMSATPEIQLAEFLKSYLVLEAAGOV  
 GACRDVQQLVQVQFSGKREKVEILRNIGDERTVVFETKKADFTAFQCKEIS  
 TTSIHDEQREORALGDFPRFGKCPVLVATSVARGLDIENVQHTINDLSTIDEY  
 VHRIGTRCGNTRALSPFDESNDLQPLVYKLTDAQODVPALVEIARSTYIPG  
 FSGTRGVPSVDRKXSLTNTAGSSSRANPVDSDSD"

## ORIGIN

Query Match 100.0%; Score 2224; DB 9; Length 2224;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2224; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ACTGAAGTGCACATGGGGGATGAAGATGGGAAGCAGAAATACACCTCATATGCTTC 60  
 1 ACTGAAGTGCACATGGGGGATGAAGATGGGAAGCAGAAATACACCTCATATGCTTC 60  
 61 CTATGTCCTCATATTTGAGAGATAGGTAATCTGAGAAATGAGACAAATTTAAACG 120  
 61 CTATGTCCTCATATTTGAGAGATAGGTAATCTGAGAAATGAGACAAATTTAAACG 120  
 121 GACTCAGCTCATCATAGAAATGAGATGAGACCTTCTGGAAGATCATTTCTGAA 180  
 121 GACTCAGCTCATCATAGAAATGAGATGAGACCTTCTGGAAGATCATTTCTGAA 180  
 181 AAGTGATTTGCTCTGAGCGGAAATTTTGGAAAACAGAGATGCTGAGTGAATTAACG 240  
 181 AAGTGATTTGCTCTGAGCGGAAATTTTGGAAAACAGAGATGCTGAGTGAATTAACG 240  
 181 AAGTGATTTGCTCTGAGCGGAAATTTTGGAAAACAGAGATGCTGAGTGAATTAACG 240  
 241 AGATTAATACATCCCAATGGGTGTTTGGAGTTGGAAGAAGTTTGGAAAACAGAGATT 300  
 241 AGATTAATACATCCCAATGGGTGTTTGGAGTTGGAAGAAGTTTGGAAAACAGAGATT 300  
 301 TTCAAAACGAGGTTTGAAGATGAGTAAGCTCTGCTTTCTGAGAGAGCTTAATGA 360  
 301 TTCAAAACGAGGTTTGAAGATGAGTAAGCTCTGCTTTCTGAGAGAGCTTAATGA 360  
 361 CTGGAAGATTAATCCCAACGGAACAGAGGTTTTCACAGAGAGGCGCTATCCGAGTGG 420  
 361 CTGGAAGATTAATCCCAACGGAACAGAGGTTTTCACAGAGAGGCGCTATCCGAGTGG 420  
 421 AAATTAATTCAGAACTTCAGGGCCATACAGAAAGGTGAGAGAGTATTCGAGGTTG 480  
 421 AAATTAATTCAGAACTTCAGGGCCATACAGAAAGGTGAGAGAGTATTCGAGGTTG 480

421 AAATTAATTCAGAACTTCAGGGCCATACAGAAAGGTGAGAGAGTATTCGAGGTTG 480  
 481 CCGTGAAGATTTGGTTAGAAAGTCCAAATTAATGCTTAGACCCAGACGAATGTATGA 540  
 481 CCGTGAAGATTTGGTTAGAAAGTCCAAATTAATGCTTAGACCCAGACGAATGTATGA 540  
 541 GCGCACTGTGGCTTTTGTCTTGAAGACCAATTAATGCTGACACGATATGTGTA 600  
 541 GCGCACTGTGGCTTTTGTCTTGAAGACCAATTAATGCTGACACGATATGTGTA 600  
 601 TACTTCCAAAGCAAGAGTGGACAGTGAACAGAGTGGTTACAAAGTTTAAATGA 660  
 601 TACTTCCAAAGCAAGAGTGGACAGTGAACAGAGTGGTTACAAAGTTTAAATGA 660  
 661 AGAAGTAAATACAGGCTCTGAAAAGAAATCTTGGAAATGAGAGCAAGAGAGAAAG 720  
 661 AGAAGTAAATACAGGCTCTGAAAAGAAATCTTGGAAATGAGAGCAAGAGAGAAAG 720  
 721 TAGTATAGTCAAGGACCAAAAGTGAACCTACATACCCCTCTCCACCTGAGATGAGA 780  
 721 TAGTATAGTCAAGGACCAAAAGTGAACCTACATACCCCTCTCCACCTGAGATGAGA 780  
 781 CTCCATCTTTCACATTAATCAGACAGGCAATTAATCTGACAAATACGACATATCTTGT 840  
 781 CTCCATCTTTCACATTAATCAGACAGGCAATTAATCTGACAAATACGACATATCTTGT 840  
 841 GGAAGTGTCTGGAATGATGACACACAGCAATTCGACTTTTGAAGAGCTTAATCTGT 900  
 841 GGAAGTGTCTGGAATGATGACACACAGCAATTCGACTTTTGAAGAGCTTAATCTGT 900  
 901 TGACACCTGATTAACACATTTGCTTAAGCTGTTTACTAAGCTTACTCTGTGCAAAA 960  
 901 TGACACCTGATTAACACATTTGCTTAAGCTGTTTACTAAGCTTACTCTGTGCAAAA 960  
 961 ATACAGATTCCTCATATCAATTCAGACAGAGATTTGATGCTGTGCTCAACAGAGGTC 1020  
 961 ATACAGATTCCTCATATCAATTCAGACAGAGATTTGATGCTGTGCTCAACAGAGGTC 1020  
 1021 TGGGAAGACCTGCGGCTTTTCTCTCAATTTTGGCTCATATGATGATGATGAATAC 1080  
 1021 TGGGAAGACCTGCGGCTTTTCTCTCAATTTTGGCTCATATGATGATGATGAATAC 1080  
 1081 TGGCAGTGTCTTAAAGAGTGGCAGAAACAGAGTGAATTAATGAGCAACCACTCGAGA 1140  
 1081 TGGCAGTGTCTTAAAGAGTGGCAGAAACAGAGTGAATTAATGAGCAACCACTCGAGA 1140  
 1141 ATTGTCACACAGATTAATTTGGAAGCCAGAAATTTCTTTTGGACCTTGTGAAGGC 1200  
 1141 ATTGTCACACAGATTAATTTGGAAGCCAGAAATTTCTTTTGGACCTTGTGAAGGC 1200  
 1201 TGTGTTATATATGAGGGAACCCAGCTGGGACATTCATTCAGCAATTAATGATGATGAT 1260  
 1201 TGTGTTATATATGAGGGAACCCAGCTGGGACATTCATTCAGCAATTAATGATGATGAT 1260  
 1261 TAAATATATATGAGTACTCTGGAAGACTGATGATATATGATGATGATGATGATGAT 1320  
 1261 TAAATATATATGAGTACTCTGGAAGACTGATGATATATGATGATGATGATGATGAT 1320  
 1321 TCTCAAAACAGATTAATTTGTTTGAATGATGATGATGATGATGATGATGATGATGAT 1380  
 1321 TCTCAAAACAGATTAATTTGTTTGAATGATGATGATGATGATGATGATGATGATGAT 1380  
 1381 TGGTCAGAAATGAGAGATTAATTTCTTGGCAGAAATGCCATCAAAAGGACAGGCGCA 1440  
 1381 TGGTCAGAAATGAGAGATTAATTTCTTGGCAGAAATGCCATCAAAAGGACAGGCGCA 1440  
 1441 AACCTTAATGTCAGTGAACCTTTCCAGAGAAATTCAAAGGTTGGCTGACAGATTTT 1500  
 1441 AACCTTAATGTCAGTGAACCTTTCCAGAGAAATTCAAAGGTTGGCTGACAGATTTT 1500  
 1501 AAAGTCAATTAATCTGTTTGTGCTGTTGACAAAGTGGTGAAGCATGATGATGATCA 1560  
 1501 AAAGTCAATTAATCTGTTTGTGCTGTTGACAAAGTGGTGAAGCATGATGATGATCA 1560







Query Match: 89.8%; Score 1997.2; DB 9; Length 2263;  
Best Local Similarity 95.1%; Pred. No. 0;  
Matches 211; Conservative 8; Indels 102; Gaps 1

D	b	25	ACTTGAAGCCACATGGAGAGATGAGATGGAGAGCAAAATCAACCTCATATGCTTC	84
Q	y	61	CTAATTTCCCATATTGTGAGAGATAGATTTCTGGAGAAATGGAGCAATTTTAACG	120
D	b	85	CTATGTTCCCATTTATGAGAGATAGATTTCTGGAGAAATGGAGCAATTTTAACG	144
Q	y	121	GACTCCAGCTTCATCATCAGAAATGATGATGACCTTCCAGAGATCAATTCATGAA	180
D	b	145	GACTCCAGCTTCATCATCAGAAATGATGATGACCTTCCAGAGATCAATTCATGAA	204
Q	y	181	AAGTGAATTTCCCTCGGGCGGAATTTTGGAAAACAGAGTGGTGGTGTGATTAAGC	240
D	b	205	AAGTGAATTTCCCTCGGGCGGAATTTTGGAAAACAGAGTGGTGGTGTGATTAAGC	264
Q	y	241	AGATATATCATCCAAATGGGTGTTTGGAGTTGAAAGATTTTGGAAAACAGAGTTT	300
D	b	265	AGATATATCATCCAAATGGGTGTTTGGAGTTGAAAGATTTTGGAAAACAGAGTTT	324
Q	y	301	TTCAACAGCAGGCTTGAATGATGATAGCTCGTTTCTGGAAGAGCTCTAGTAATGA	360
D	b	325	TTCAACAGCAGGCTTGAATGATGATAGCTCGTTTCTGGAAGAGCTCTAGTAATGA	384
Q	y	361	CTGCGAAGATATATCCAAACAGGAAACAGAGGTTTTCAGAGAGCGGCTATCGAGATGG	420
D	b	385	CTGCGAAGATATATCCAAACAGGAAACAGAGGTTTTCAGAGAGAGCGG-----	431
Q	y	421	AAATATTTCAAAACCTTCAGGGCCATACAGAAAGGTGAAAGGTATTTCCAGAGTGG	480
D	b	432	-----	431
Q	y	481	CCGTGAGAGATTTGCTCTAGAAAGTCCAAATAATGACTTGAACCCAGACGATGATGCA	540
D	b	432	-----ATATGACTTGAACCCAGACGATGATGCA	462
Q	y	541	GCGCAGCTGGTGGCTTTTGGTTCTTGAAGAACAGATTAATGATGCAACAGGATATGGTGA	600
D	b	463	GCGCAGCTGGTGGCTTTTGGTTCTTGAAGAACAGATTAATGATGCAACAGGATATGGTGA	522
Q	y	601	TACTTCTCAAAAGCAGAAAGTGACAGTGGAAATGCAAGGTGTTTAATGA	660
D	b	523	TACTTCTCAAAAGCAGAAAGTGACAGTGGAAATGCAAGGTGTTTAATGA	582
Q	y	661	AGAAATATAACAGGCTCTGGAAAGAAATCTTGGAAAGTGAAGAGAGAGAGAGAAAG	720
D	b	583	AGAAATATAACAGGCTCTGGAAAGAAATCTTGGAAAGTGAAGAGAGAGAGAGAAAG	642
Q	y	721	TAGTATATCTCAAGAACAAAGTGAACCTACATACCCCTCTCCACTGAGATGAGA	780
D	b	643	TAGTATATCTCAAGAACAAAGTGAACCTACATACCCCTCTCCACTGAGATGAGA	702
Q	y	781	CTCCATCTTTTCCCATATATGAGACAGGCAATTAATCTTGAACCAATGATCTTGT	840
D	b	703	CTCCATCTTTTCCCATATATGAGACAGGCAATTAATCTTGAACCAATGATCTTGT	762
Q	y	841	GGAAAGTGTCTGACATGATGACACACACAGCAATCTGACTTTGAAAGACTAATCTGTG	900
D	b	763	GGAAAGTGTCTGACATGATGACACACACAGCAATCTGACTTTGAAAGACTAATCTGTG	822
Q	y	901	TCAGACACTGAATTAACACAACTGCTTAAAGTGGTTTACTAGCTTATCCCGTGCAAA	960
D	b	823	TCAGACACTGAATTAACACAACTGCTTAAAGTGGTTTACTAGCTTATCCCGTGCAAA	882
Q	y	961	ATACAGATTTCTATCACTCTGACAGACAGAGATTTGATGCTTGTCTCAACAGGGTC	1020
D	b	883	ATACAGATTTCTATCACTCTGACAGACAGAGATTTGATGCTTGTGTCTCAACAGGGTC	942
Q	y	1021	TGGAAAGCTCGGCTTTTCTCCCAACAAATTTGGTCTCATATGATGATGATGAAATAC	1080
D	b	943	TGGAAAGCTCGGCTTTTCTCCCAACAAATTTGGTCTCATATGATGATGATGAAATAC	1002
Q	y	1081	TGCCAGTGTGTTTAAAGAGTTGACAGAACAGAGTGTATTAATTTGACACCACTCGAGA	1140

Db 1003 TGCCAGTCGTTTAAAGGTTGCGAGAACGAGAGTATATTGTAGCACCAACTGGA 1062  
 Qy 1141 ATTGCTCAACGAGTTTATTTTGAAGCCAGAAAAATTTCTTTGGGACTTGTGTAAGAC 1200  
 Db 1063 ATTGCTCAACGAGTTTATTTTGAAGCCAGAAAAATTTCTTTGGGACTTGTGTAAGAC 1122  
 Qy 1201 TGTGTATATATGAGGAGAACCCAGCTGGGACATTCATTCGCAATAGTACAGAGCTG 1260  
 Db 1123 TGTGTATATATGAGGAGAACCCAGCTGGGACATTCATTCGCAATAGTACAGAGCTG 1182  
 Qy 1261 TATATATATATGAGTACTCTCTGGAAGAAGTATGATATATAGGCAAGAAAAAGATTG 1320  
 Db 1183 TATATATATATGAGTACTCTCTGGAAGAAGTATGATATATAGGCAAGAAAAAGATTG 1242  
 Qy 1321 TCTCAACAGATCAAAATCTAGTTTGGATGAGATGAGATGAGATGAGATGAGATG 1380  
 Db 1243 TCTCAACAGATCAAAATCTAGTTTGGATGAGATGAGATGAGATGAGATGAGATG 1302  
 Qy 1381 TGTTCAGAAATGAAGAGATTATTTCTTCCAGAGAAATGCAATCAAGAGACAGCCCA 1440  
 Db 1303 TGTTCAGAAATGAAGAGATTATTTCTTCCAGAGAAATGCAATCAAGAGACAGCCCA 1362  
 Qy 1441 AACCTTATGTCAGTGAATTTTCCAGAGAAATCAAGAGTGGCTGACAGTTT 1500  
 Db 1363 AACCTTATGTCAGTGAATTTTCCAGAGAAATCAAGAGTGGCTGACAGTTT 1422  
 Qy 1501 AAAGTCAATATATCTGTTGTTGCTGTGGAAGAGTGGTGAAGATGAGATGATG 1560  
 Db 1423 AAAGTCAATATATCTGTTGTTGCTGTGGAAGAGTGGTGAAGATGAGATGATG 1482  
 Qy 1561 GAGAGCGTGTCCAGAGTGGCCAGTGTCTCAAGAAAGAGAAAGTGGTGAATGATG 1620  
 Db 1483 GAGAGCGTGTCCAGAGTGGCCAGTGTCTCAAGAAAGAGAAAGTGGTGAATGATG 1542  
 Qy 1621 AAACATAGGGAGTGAAGAACTATGCTTTGTTGAACTPAAGAAAAAGCAGATTTC 1680  
 Db 1543 AAACATAGGGAGTGAAGAACTATGCTTTGTTGAACTPAAGAAAAAGCAGATTTC 1602  
 Qy 1681 TGCAACTTTCTTTGTCAAGAAAAATATCAATACAGAGATGCAATGATGCGGAA 1740  
 Db 1603 TGCAACTTTCTTTGTCAAGAAAAATATCAATACAGAGATGCAATGATGCGGAA 1662  
 Qy 1741 GAGAGAGCGGAGAGAGCTCTTGAGAGATTTGCTTGGTGAAGTGGCCAGTTCTTTC 1800  
 Db 1663 GAGAGAGCGGAGAGAGCTCTTGAGAGATTTGCTTGGTGAAGTGGCCAGTTCTTTC 1722  
 Qy 1801 TACTTCAGTACGTCGAGAGAGGCTGATATGGAAGATGTCACATGTTATCAATTTGA 1860  
 Db 1723 TACTTCAGTACGTCGAGAGAGGCTGATATGGAAGATGTCACATGTTATCAATTTGA 1782  
 Qy 1861 TCTTCTCTTACCAATGATGATATGTTCAATGCAATTGGGCGTACTGTTGTTGGAA 1920  
 Db 1783 TCTTCTCTTACCAATGATGATATGTTCAATGCAATTGGGCGTACTGTTGTTGGAA 1842  
 Qy 1921 TACTGGCAGAGCAATTTCTTTTGTGATCTTGAATGGATTAACATTTTGAACAGCTCT 1980  
 Db 1843 TACTGGCAGAGCAATTTCTTTTGTGATCTTGAATGGATTAACATTTTGAACAGCTCT 1902  
 Qy 1961 AGTAAAGATATGACAGATGCTCAACAGAGATGTTCTCTGCAATGTTGGAAGAAATTG 2040  
 Db 1903 AGTAAAGATATGACAGATGCTCAACAGAGATGTTCTCTGCAATGTTGGAAGAAATTG 1962  
 Qy 2041 TAGATCATATGATCTGCTGCTCAGTGTGATGACAGAGAGAAAGCTGTTTGCATGATTA 2100  
 Db 1963 TAGATCATATGATCTGCTGCTCAGTGTGATGACAGAGAGAAAGCTGTTTGCATGATTA 2022  
 Qy 2101 TACAGAAAGGCAAGAGCACTTTGAACACAGCTGGGTTTCTTCTTCAAGAGCTCCAA 2160  
 Db 2023 TACAGAAAGGCAAGAGCACTTTGAACACAGCTGGGTTTCTTCTTCAAGAGCTCCAA 2082  
 Qy 2161 TCCAGTATGATATGATGATGAGATTAAGAGCAAAATATCTCAGTGTGCTGTTTGA 2220  
 Db 2083 TCCAGTATGATATGATGATGAGATTAAGAGCAAAATATCTCAGTGTGCTGTTTGA 2142

Qy 2221 TGCA 2224  
 Db 2143 TGCA 2146  
 RESULT 4  
 HSM802178  
 LOCUS HSM802178 2159 bp mRNA linear PRI 18-FEB-2000  
 DEFINITION Homo sapiens mRNA; cDNA DKFp434B1122 (from clone DKFp434B1122);  
 partial cds.  
 VERSION AL137462  
 KEYWORDS AL137462.1 GI:6808043  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 2189)  
 AUTHORS Blum, H., Bauerhach, S., Mewes, H.W., Gassenhuber, J. and Wiemann, S.  
 TITLE Direct Submission  
 JOURNAL Submitted (15-JAN-2000) MIPS, Am Klopferplatz 18a, D-82152  
 Martinsried, GERMANY  
 COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
 Research Center (DKFZ), Email: s.wiemann@kfz-heidelberg.de;  
 sequenced by LMC (Lucwig Maximilians University,  
 Munich/Germany) within the cDNA sequencing consortium of the German  
 Genome Project.  
 This clone (DKFp434B1122) is available at the RZPD in Berlin.  
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further  
 information about the clone and the sequencing project is available  
 at <http://www.mips.biochem.mpg.de/proj/cDNA/>.  
 Location/Qualifiers  
 1..2189  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="DKFp434B1122"  
 /tissue\_type="testis"  
 /clone\_1db="434 (synonym: hncs3). Vector pSport1; host  
 DH10B; sites NotI + SalI"  
 /dev\_stage="adult"  
 1..2189  
 /gene="DKFp434B1122"  
 <1..1910  
 /gene="DKFp434B1122"  
 /note="strong similarity to testis specific Helicase (VASA  
 homolog)"  
 /codon\_start=3  
 /product="hypothetical protein"  
 /protein\_id="CAB70750.1"  
 /db\_xref="GI:6808044"  
 /db\_xref="GOA:Q9NO10"  
 /db\_xref="SWISS-PROT:Q9NO10"  
 /translation="SRGNRCFSNSRPFDDSGCFRSESSNDCENPFRNRFSGRGY  
 RDGNSEFAGPRYRGRGSGFRRCRGFGIAGNNLDPDECMOPGGLFGSRRLVLSG  
 TNGDTSQSRSRGSRGSKLNEVITGSRKRWKSRABESSDYOGRXYIIP  
 PPDEDSDIFAHVQTGINDKXDTLLIVEVSGDAPALITFEANLCTLNNTAKAG  
 YTKLTPVQXYSIPILLAGRDLMACAQDSGKTAFLPILAHMEDIGTLASRFKELQ  
 PECIIVAPRELWQIYLARAFSGTCVRAVYAGTQLGHSISQIIVOGNNICATP  
 GRMLDIIGKEKIGLKOIKYLVDSEADRMIDMGFGEMKLLISCPMPSEKORULMF  
 ATPPEIOLLAELFKSNLFPVAVGVGAGACDVOQVLYOVGFSEKELVILNIG  
 DERTWVFEVTKKKADRIATFLGQEKISTTSJHGDESRERBOALGDPFGCPVLVAT  
 SVAAKSLDIENQHYINFLPSTIDBYHRIKRTKRCNTRGALISFPLSDNHLAOP  
 LVKVLTDAGQDDVPAMWLEETAFSTYIPGSGSIRGNVFAVSDTRKSKSLNTNAGSSSQ  
 APEPVDSQSD"  
 2128..2133  
 /gene="DKFp434B1122"  
 2147  
 /gene="DKFp434B1122"  
 polyA\_signal  
 polyA\_site  
 ORIGIN

Query Match		87.2%:	Score 1939.6:	DB 9:	Length 2189:
Beet Local Similarity		99.8%:	Pred. No. 0:		
Matches 1942:		Conservative 0:	Mismatches 4:	Indels 0:	Gaps 0:
QY	279	AGAGTTTGGAAAAGAGTTTTCACACAGAGTTTGAAGATGTGATAGCTGTGTT	338		
DB	1	AGAGTTTGGAAAAGAGTTTTCACACAGAGTTTGAAGATGTGATAGCTGTGTT	60		
QY	339	TCTGGAGAGATCTAGATATGATCTGCGAAGATATCCACACGGAACAGAGGTTTCCA	398		
DB	61	TCTGGAGAGATCTAGATATGATCTGCGAAGATATCCACACGGAACAGAGGTTTCCA	120		
QY	399	AGAGAGCGGCTATCGAGATGAGAAATATTCAGAGGCTTCAGGCGCATACAGAGAGTG	458		
DB	121	AGAGAGCGGCTATCGAGATGAGAAATATTCAGAGGCTTCAGGCGCATACAGAGAGTG	180		
QY	459	GAAAGGTAGTTTCCGAGGTTGCGGTGAGAGATTTGGTCTAGAGAACTCCAAATATGACT	518		
DB	181	GAAAGGTAGTTTCCGAGGTTGCGGTGAGAGATTTGGTCTAGAGAACTCCAAATATGACT	240		
QY	519	TAGACCCAGACGAATGTATGACGCGCATGCTGCGCTTTTGGTCTAGAGAACCCAGTAT	578		
DB	241	TAGACCCAGACGAATGTATGACGCGCATGCTGCGCTTTTGGTCTAGAGAACCCAGTAT	300		
QY	579	TAACTGGACAGATATATGTATCTCTCAGAGCAAGATGCGAGTGAAGTGAACGAG	638		
DB	301	TAACTGGACAGATATATGTATCTCTCAGAGCAAGATGCGAGTGAAGTGAACGAG	360		
QY	639	GTGGTTACAAAGTTTAAATGAAGATATTAACAGGCTCTGGAAGAAATCTTGAAGT	698		
DB	361	GTGGTTACAAAGTTTAAATGAAGATATTAACAGGCTCTGGAAGAAATCTTGAAGT	420		
QY	699	CAGAGAGCAAGAGAGAGAGATGTATCTCAGAGCAAGATGCGAGTGAAGTGAACGAG	758		
DB	421	CAGAGAGCAAGAGAGAGAGATGTATCTCAGAGCAAGATGCGAGTGAAGTGAACGAG	480		
QY	759	CTCTCCACCTGAGAGATGAGATCTCATCTTTCGACATTAATCAGACAGGATTAAGTCTG	818		
DB	481	CTCTCCACCTGAGAGATGAGATCTCATCTTTCGACATTAATCAGACAGGATTAAGTCTG	540		
QY	819	ACAAATACGACATTAATCTTGTGGAAGTGTCTGACATGATGACCAACGACATTTCTGA	878		
DB	541	ACAAATACGACATTAATCTTGTGGAAGTGTCTGACATGATGACCAACGACATTTCTGA	600		
QY	879	CTTTTGAAGAAAGTATCTCTGTCAGACATGATTAACACATTTGCTAAAGCTGTTATA	938		
DB	601	CTTTTGAAGAAAGTATCTCTGTCAGACATGATTAACACATTTGCTAAAGCTGTTATA	660		
QY	939	CTAAAGCTTACTCTCTGTCGAAAAATACGATTTCTCATACTTTCGAGAGCGAGATTTGA	998		
DB	661	CTAAAGCTTACTCTCTGTCGAAAAATACGATTTCTCATACTTTCGAGAGCGAGATTTGA	720		
QY	999	TGGCTTGTGCTCAACAGGCTCTGGAAGATGCGGCTTTTCTCTACCAATTTTGGCTC	1058		
DB	721	TGGCTTGTGCTCAACAGGCTCTGGAAGATGCGGCTTTTCTCTACCAATTTTGGCTC	780		
QY	1059	ATAATGATGATATGGAATTAATCTGCGAGTCTGTTTAAAGATTTGACAGAACGAGATGTA	1118		
DB	781	ATAATGATGATATGGAATTAATCTGCGAGTCTGTTTAAAGATTTGACAGAACGAGATGTA	840		
QY	1119	TTATTTGAGACCAACTCGAGAAATTTGCTCAACAGATTTTATTTGGAAGCAGAAATTTT	1178		
DB	841	TTATTTGAGACCAACTCGAGAAATTTGCTCAACAGATTTTATTTGGAAGCAGAAATTTT	900		
QY	1179	CTTTTGGAGCTTGTGTAAAGCTGTGTATATATGCGGGAACCCAGCTGGACATTTCAA	1238		
DB	901	CTTTTGGAGCTTGTGTAAAGCTGTGTATATATGCGGGAACCCAGCTGGACATTTCAA	960		
QY	1239	TTTGCACAAATATGATCAAGGCTGTAAATATATATATGCTGTAATCTCTGGAAGCTGATGATA	1298		
DB	961	TTTGCACAAATATGATCAAGGCTGTAAATATATATATGCTGTAATCTCTGGAAGCTGATGATA	1020		
QY	1299	TCATATGCGAAAGAAAGATTTGCTCAACAGATCAAAATCTTAGTTTGGATGAAGCTG	1358		

DB	1021	TCATATGCGAAAGAAAGATTTGCTCAACAGATCAAAATCTTAGTTTGGATGAAGCTG	1080		
QY	1359	ATCCGATGTTGATATGATGTTTGGTCCAGAAATGAAGAAATTAATTTCTTGCCAGGAA	1418		
DB	1081	ATCCGATGTTGATATGATGTTTGGTCCAGAAATGAAGAAATTAATTTCTTGCCAGGAA	1140		
QY	1419	TGCGATCAAGAGAAACAGCGGCAACCTTATGTTTCAGTGCATTTTCCAGAGAAATTC	1478		
DB	1141	TGCGATCAAGAGAAACAGCGGCAACCTTATGTTTCAGTGCATTTTCCAGAGAAATTC	1200		
QY	1479	AAAGGTCGTGAGAAATTTTAAAGTCAATATCTGTTTGTGCTGTGCAAGATG	1538		
DB	1201	AAAGGTCGTGAGAAATTTTAAAGTCAATATCTGTTTGTGCTGTGCAAGATG	1260		
QY	1539	GTGAGATATGATGATGTTTCAAGACGCTTCTCAAGTTGGCCAGTTCTCAAAAGAG	1598		
DB	1261	GTGAGATATGATGATGTTTCAAGACGCTTCTCAAGTTGGCCAGTTCTCAAAAGAG	1320		
QY	1599	AAAGGTCGTGAGAAATTTCTGCAAGATGAGGAGTAAAGAACTATGCTTTTGTGAAA	1658		
DB	1321	AAAGGTCGTGAGAAATTTCTGCAAGATGAGGAGTAAAGAACTATGCTTTTGTGAAA	1380		
QY	1659	CTAAGAAAAAGCAGATTTTACTGCACTTTTCTTGTCAAGAAAAATATCACTACAA	1718		
DB	1381	CTAAGAAAAAGCAGATTTTACTGCACTTTTCTTGTCAAGAAAAATATCACTACAA	1440		
QY	1719	GTATCATGTTGATCCGGAACAGAGAGCGGAGCAAGCTTTGAGAGATTTTCCGCTTG	1778		
DB	1441	GTATCATGTTGATCCGGAACAGAGAGCGGAGCAAGCTTTGAGAGATTTTCCGCTTG	1500		
QY	1779	GAAGTGGCCAGATTTCTTGTCTACTTCACTACTGCGCAGAGGCTGCAATTTGAAAATG	1838		
DB	1501	GAAGTGGCCAGATTTCTTGTCTACTTCACTACTGCGCAGAGGCTGCAATTTGAAAATG	1560		
QY	1839	TGCAACATGTTATCAATTTTATGATCTTCTCTACATTAATGATATGATCAATGCAATG	1898		
DB	1561	TGCAACATGTTATCAATTTTATGATCTTCTCTACATTAATGATATGATCAATGCAATG	1620		
QY	1899	GCGGTATGCTGCTGTTGGGGAATATCTGCAAGACAAATTTCTTTTGTGATCTTGAATCG	1958		
DB	1621	GCGGTATGCTGCTGTTGGGGAATATCTGCAAGACAAATTTCTTTTGTGATCTTGAATCG	1680		
QY	1959	ATAACATTTAGACAGGCTCTAGTAAAGATTTGACAGATGCTCAACAGATGTTCTCTG	2018		
DB	1681	ATAACATTTAGACAGGCTCTAGTAAAGATTTGACAGATGCTCAACAGATGTTCTCTG	1740		
QY	2019	CATGCTTGGAGAAATTTGCTTTAGTACATCACTTCTGCTTCACTGCTGATGACAGAG	2078		
DB	1741	CATGCTTGGAGAAATTTGCTTTAGTACATCACTTCTGCTTCACTGCTGATGACAGAG	1800		
QY	2079	GAAAGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	2138		
DB	1801	GAAAGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1860		
QY	2139	TTTCTTTCTTCAAGATCTCCATTCAGATGATGATGATGATGATGATGATGATGATGATG	2198		
DB	1861	TTTCTTTCTTCAAGATCTCCATTCAGATGATGATGATGATGATGATGATGATGATGATG	1920		
QY	2199	TCTTCAAGCTGCTGTTTGTGATGCA	2224		
DB	1921	TCTTCAAGCTGCTGTTTGTGATGCA	1946		

RESULT 5  
AX748127 1984 bp mRNA 1linear PART 20-JUN-2003  
LOCUS AX748127  
DEFINITION Sequence 1652 from Patent EP1308459.  
ACCESSION AX748127  
VERSION AX748127.1 GI:32132515  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

ACCESION	TO DEAD BOX PROTEIN 4.
AK093439	
AK093439.1	GI:21752310
VERSION	

RESULT 6	AK093439
LOCUS	AK093439
DEFINITION	AK093439 Homo sapiens CDNA FLJ36120 f15, clone TEST2022736, highly similar
ACCESSION	AK093439
VERSION	AK093439.1 GI:21752310



KEYWORDS oligo capping; file (full insert sequence).  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE  
AUTHORS  
1 Nishi, T., Ota, T., Nakagawa, S., Senoh, A., Mizuguchi, H., Inagaki, H., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Makatsuki, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sakine, M., Kikuchi, H., Kanda, K., Magatsuma, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Negishi, K., Masuno, Y., Nagai, K. and Isozaki, T.  
NEO human cDNA sequencing project  
Unpublished  
2 (bases 1 to 1984)  
Isozaki, T. and Yamamoto, J.  
Direct Submission  
Submitted (04-JUL-2002) Takao Isozaki, FLJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 282-0812, Japan  
E-mail: genomics@nt.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986  
NEO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.  
Location/Qualifiers  
FEATURES  
source  
1..1984  
/organism="Homo sapiens"  
/mol\_type="RNA"  
/db\_xref="taxon:9606"  
/clone="TEST12022738"  
/issue\_type="testis"  
/clone\_id="TEST12"  
/note="Cloning vector: pME18SFL3"  
ORIGIN  
Query Match 72.0%; Score 1601; DB 9; Length 1984;  
Best Local Similarity 96.9%; Pred. No. 0;  
Matches 1662; Conservative 0; Mismatches 5; Indels 48; Gaps 1;  
QY 510 ATATGACTTAGACCCAGACAGATGTATGACGCCACTGTGCGCTTTTGGTTCTAGAA 569  
DB 117 ATATGACTTAGACCCAGACAGATGTATGACGCCACTGTGCGCTTTTGGTTCTAGAA 176  
QY 570 GACCAAGATTAAAGTGGCAAGGTATGTGATCTTCTCAAGCAAGAGTGGCAGTGGAA 629  
DB 177 GACCAAGATTAAAGTGGCAAGGTATGTGATCTTCTCAAGCAAGAGTGGCAGTGGAA 236  
QY 630 GTGAACGAGTGGTTACAAAGTTTAATGAAGAATATACAGGCTCTGGAAGAATT 689  
DB 237 GTGAACGAGTGGTTACAAAGTTTAATGAAGAATATACAGGCTCTGGAAGAATT 248  
QY 690 CTGGAAGTCAAGACAGAGAGAGAGAAAGTATGATCTCAAGACCAAAAGTACCT 749  
DB 249 CTGGAAGTCAAGACAGAGAGAGAGAAAGTATGATCTCAAGACCAAAAGTACCT 308  
QY 750 ACATACCCCTCTCTCCACCTGAGTGAAGTCCATCTTGGACATTATCAGACAGCA 809  
DB 309 ACATACCCCTCTCTCCACCTGAGTGAAGTCCATCTTGGACATTATCAGACAGCA 368  
QY 810 TAACTTCGACAATATGACACTATCTTGTGGAAGTCTGACATGATGACACACAG 869  
DB 369 TAACTTCGACAATATGACACTATCTTGTGGAAGTCTGACATGATGACACACAG 428  
QY 870 CAATTTCGACTTTTGAAGAAGTATCTCTGTCACACATGAATTAACAACATTGCTAAG 929  
DB 429 CAATTTCGACTTTTGAAGAAGTATCTCTGTCACACATGAATTAACAACATTGCTAAG 488  
QY 930 CTGGTATATCTAGCTACTCTGTGCAAAATATACGATTCCTATCTTGCAGGAC 989

DB 489 CTGGTATATCTAGCTACTCTGTGCAAAATATACGATTCCTATCTTGCAGGAC 548  
QY 990 GAGATTGATGAGCTTTGGCTTCAACAGGCTGTGGAGACGTGGGCTTTCTCTACCA 1049  
DB 549 GAGATTGATGAGCTTTGGCTTCAACAGGCTGTGGAGACGTGGGCTTTCTCTACCA 608  
QY 1050 TTTGGCTCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1109  
DB 609 TTTGGCTCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 668  
QY 1110 CAGAGTATATTTGAGACCAACTGACATGATGATGATGATGATGATGATGATGATG 1169  
DB 669 CAGAGTATATTTGAGACCAACTGACATGATGATGATGATGATGATGATGATGATG 728  
QY 1170 GAAATTTTCTTTGGGAGCTGTGTAAGAGCTGTGTAATATATATGAGGGAACCCAGCTGG 1229  
DB 729 GAAATTTTCTTTGGGAGCTGTGTAAGAGCTGTGTAATATATGAGGGAACCCAGCTGG 788  
QY 1230 GACATTCAATTCGACAAATATGATGATGATGATGATGATGATGATGATGATGATG 1289  
DB 789 GACATTCAATTCGACAAATATGATGATGATGATGATGATGATGATGATGATGATG 848  
QY 1290 TGATGATATCATAGGCAAGAAAGATGATGATGATGATGATGATGATGATGATGATG 1349  
DB 849 TGATGATATCATAGGCAAGAAAGATGATGATGATGATGATGATGATGATGATGATG 908  
QY 1350 ATGAGCTGATGCGATGTTGGATGATGATGATGATGATGATGATGATGATGATGATG 1409  
DB 909 ATGAGCTGATGCGATGTTGGATGATGATGATGATGATGATGATGATGATGATGATG 968  
QY 1410 GCCCAAGATGCCATCAAGAAAGAAAGAGGCGCAACCCCTATATGTCAGTCACTTTCCAG 1469  
DB 969 GCCCAAGATGCCATCAAGAAAGAAAGAGGCGCAACCCCTATATGTCAGTCACTTTCCAG 1028  
QY 1470 AGGAATTCGAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1529  
DB 1029 AGGAATTCGAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1088  
QY 1530 GACAAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1589  
DB 1089 GACAAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1148  
QY 1590 CAAAAAGAAAAGCTCGTTGAATTTGCGAAACATAGGCGATGAAAGAACTATAGTCT 1649  
DB 1149 CAAAAAGAAAAGCTCGTTGAATTTGCGAAACATAGGCGATGAAAGAACTATAGTCT 1208  
QY 1650 TTGTTGAACTAAGAAAAGAGATTTTACTGCAACTTTCTTGTCAAGAAAATAT 1709  
DB 1209 TTGTTGAACTAAGAAAAGAGATTTTACTGCAACTTTCTTGTCAAGAAAATAT 1268  
QY 1710 CAACATCAAGTATCATGTGTATGCGGAAACAGAGAGAGGAGAGAGCTTTGGAGAT 1769  
DB 1269 CAACATCAAGTATCATGTGTATGCGGAAACAGAGAGAGGAGAGAGCTTTGGAGAT 1328  
QY 1770 TTGCTTTGGAAGAGTCCAGTCTTGTGCTACTCAATGATGCGAGAGGCTGGATA 1829  
DB 1329 TTGCTTTGGAAGAGTCCAGTCTTGTGCTACTCAATGATGCGAGAGGCTGGATA 1388  
QY 1830 TTGAAATGTGCAACATGTTATCAATTTTGATTTTCTTCTACATGATGAAATATGTC 1889  
DB 1389 TTGAAATGTGCAACATGTTATCAATTTTGATTTTCTTCTACATGATGAAATATGTC 1448  
QY 1890 ATGGAATGGGCTATCGGTCGTTGGGAATCTGCGAGAGCAATTTCTTTTGTATC 1949  
DB 1449 ATGGAATGGGCTATCGGTCGTTGGGAATCTGCGAGAGCAATTTCTTTTGTATC 1508  
QY 1950 TTGAATCGATTAACCAATTTAGCACAGCTTCAATTAAGTATGACAGATCTCAACAG 2009  
DB 1509 TTGAATCGATTAACCAATTTAGCACAGCTTCAATTAAGTATGACAGATCTCAACAG 1568  
QY 2010 ATGTTCTGATGATGTTGGAAGAAATGCTTTTGTATCATGATCTCTGAGTGGTA 2069



Db 1569 ATGTCCTGCATGGTGGAGAAATGGCTTTAGTACATTCCTGGCTTCAGTGTGA 1628

Qy 2070 GTACAGAGAGAAAGTGTTCATCATGTTGATACAGAAAGGCGAGAGACTTTGACA 2129

Db 1629 GTACAGAGAGAAAGTGTTCATCATGTTGATACAGAAAGGCGAGAGACTTTGACA 1688

Qy 2130 CAGCTGGGTTTCTTCTTCACAGAGCTCCCAATCCAGTAGATGATGATGAGATTTAA 2189

Db 1689 CAGCTGGGTTTCTTCTTCACAGAGCTCCCAATCCAGTAGATGATGATGAGATTTAA 1748

Qy 2190 GCCAAAACATCCTCAAGTGTGTTTGTATGA 2224

Db 1749 GCCAAAACATCCTCAAGTGTGTTTGTATGA 1793

RESULT 7

S75275 3030 bp mRNA linear ROD 15-MAY-1995

LOCUS RVLG=vasa-like gene protein [rats, Wistar-Imanishi, testis, mRNA, 3030 nt].

DEFINITION S75275

VERSION S75275.1 GI:806463

KEYWORDS

SOURCE

ORGANISM

Rattus sp.

Rattus sp.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE

1 (bases 1 to 3030)

Komiyama, T. and Tanigawa, Y.

Cloning of a gene of the DEAD box protein family which is specifically expressed in germ cells in rats

Biochem. Biophys. Res. Commun. 207 (1), 405-410 (1995)

JOURNAL

MDLINE

PUBMED

7857296

Genbank staff at the National Library of Medicine created this entry [NCBI gisbseq 159937] from the original journal article.

REMARK

This sequence comes from Fig. 1.

FEATURES

source

1..3030

/organism="Rattus sp."

/mol\_type="mRNA"

/db\_xref="taxon:10118"

1..3030

/gene="RVLG"

/note="vasa-like gene protein, RVLG protein"

266..2407

/note="putative DEAD box protein; This sequence comes from Fig. 1; RVLG protein"

/gene="RVLG"

/product="vasa-like gene protein"

/protein\_id="AAB33364.1"

/db\_xref="GI:806464"

/translation="MGDEDMENALIKRVVSVVYVPEKDKTSSGANDTNNRTSASSS EYEDGSGRHDHWRSGFSGRVLGNRDISSSKRTTSTGGFGKGFGRGLANK FEEGDSGSGKSTINDCEDTQTRSGRFGKGPDPGNDSSGPRGRGRSDSEYDQC GSQGGGLFGSKRPASDSGSDTQSSGNARCAKGLNEVTVSGSKNWKSEAG GSISDIQGEKTVYIPPEDEDSIFAHYOTGINDKXDTLVISGNDAPALITTE EAMLCOTLNNNAKAGYKTKPVOKYSTIVLAGRDMACQSGTGAATLILIAH NMRDITASREKEOEPECIIVAPPRVILNOLYLAERKFSGTCTRAVYVYGTGFH STROIVOCNITICATPRGLMDIGEKIGLKVRLVDEADRLMDGFGEMKELIS CRMPSEKOROTLLSATPEELIQELAGEFLKSNVLPYAVQVGGACRVOOSLIQVG PVFKRKUVELILRNIGDERPMVFETKKKADFIATFLOEKISTTSIHGDRGERO ALGSDRCCKPVLVATSVARGLDIENOVHINFLPTIDIEYVHARGRCRGCNTER AISFDTSNDHLAOPLVKLSDAOQVPAWLERIAFSVAPEFSNSTGAVPASPDR TRKPFQGNLTNTAGTSSAQAHPNVDSDWD"

ORIGIN

Query Match 71.7%; Score 1594; DB 10; Length 3030;

Best Local Similarity 84.8%; Pred. No. 0;

Matches 1897; Conservative 0; Mismatches 270; Indels 71; Gaps 7;

Qy 1 ACTTGAGTCAACCATGGGGGATGAGATTTGGGGAAGCAAAAT---CAACCTCATATGTC 57

Db 253 ACTTGAGTCAACCATGGGAGATGAGATTTGGGGAAGCAAAATCCTCAACCTCATATGTC 312

Qy 58 TTCTATGTTCCCATATTTGGAAGAGATAGTA---TTCTGAGAGAAAATGAGACAATTT 114

Db 313 TTCTATGTTCCCATATTTGGAAGAGATAGTA---TTCTGAGAGAAAATGAGACAATTT 372

Qy 115 TTAACGAGCTCCAGTTTATCATCAGAAATGAGATGAGACTCTCTGAAAGATCATTT 174

Db 373 TTAACGAGCTCCAGTTTATCATCAGAAATGAGATGAGACTCTCTGAAAGATCATTT 432

Qy 175 CATGAAAATGAGATTTGCTCTGGGCGGAAATTTGAAAACAGAGATGCTGGTGAAGTGA 234

Db 433 CATGAAAATGAGATTTGCTCTGGGCGGAAATTTGAAAACAGAGATGCTGGTGAAGTGA 492

Qy 235 TTAAGGAGATATATCATCACAATGAGGTGTTTGGAGTTGAAAAGTTTGGAAAACAG 294

Db 433 TTAAGGAGATATATCATCACAATGAGGTGTTTGGAGTTGAAAAGTTTGGAAAACAG 552

Qy 295 AGGTTTTCACAACAGAGGTTTGAAGATGAGTGTGTTCTGAGAGAGTGTAG 354

Db 553 AGGTTTTCACAACAGAGGTTTGAAGATGAGTGTGTTCTGAGAGAGTGTAG 612

Qy 355 TTAATGACTGCGAAGATTAATCCACAGGAAACAGAGGTTTCCAGAGAGCGGCTATCG 414

Db 613 TTAATGACTGCGAAGATTAATCCACAGGAAACAGAGGTTTCCAGAGAGCGGCTATCG 672

Qy 415 AGATGGAATATATCATCAGAGGTTTCAAGGCGCATCAGAGAGGTTGAAAGAGTATTCG 474

Db 673 AGATGGAATATATCATCAGAGGTTTCAAGGCGCATCAGAGAGGTTGAAAGAGTATTCG 722

Qy 475 AGGTTGCCGTGAGAGATTTGGTCTAGAGAGTCCAAATATAGCTTAGACCCAGACGATG 534

Db 723 AGGTTGCCGTGAGAGATTTGGTCTAGAGAGTCCAAATATAGCTTAGACCCAGACGATG 747

Qy 535 TATGACAGGCACTGCTGCTTTTGGTTCTAGAAACCGATTAATAGTGCACAGATTA 594

Db 748 TATGACAGGCACTGCTGCTTTTGGTTCTAGAAACCGATTAATAGTGCACAGATTA 807

Qy 595 TGGTATATCTTCAAGAGCAAGAGTGGCAGTGAAGTGAACGAGTGTGTTTCAAGAGT 654

Db 808 TGGTATATCTTCAAGAGCAAGAGTGGCAGTGAAGTGAACGAGTGTGTTTCAAGAGT 861

Qy 655 AATGAGAGATTAATTAACAGGCTCTGGAAGAAATTTCTGGAAGTCAAGACAGAGAG 714

Db 862 AATGAGAGATTAATTAACAGGCTCTGGAAGAAATTTCTGGAAGTCAAGACAGAGAG 921

Qy 715 AGAAGTAGATATCAAGAGCAAAAGTGAACCTACATACCCCTCTCCACCTGAGGA 774

Db 922 AGAAGTAGATATCAAGAGCAAAAGTGAACCTACATACCCCTCTCCACCTGAGGA 981

Qy 775 TGAAGACTCATTTTGGACATTAATAGACAGGATTAACCTTGACAAATAGACATAT 834

Db 982 TGAAGACTCATTTTGGACATTAATAGACAGGATTAACCTTGACAAATAGACATAT 1041

Qy 835 TCTTGGAGAGTCTCGACATGATGACACACAGAAATTCGACCTTTGGAAGAACTTA 894

Db 1042 TCTTGGAGAGTCTCGACATGATGACACACAGAAATTCGACCTTTGGAAGAACTTA 1101

Qy 895 TCTTGTGACACATGAATTAACAATTTGCTAAAGTGTGTTTAAAGTCTACTCTGT 954

Db 1102 TCTTGTGACACATGAATTAACAATTTGCTAAAGTGTGTTTAAAGTCTACTCTGT 1161

Qy 955 GCAAAATACAGATTTCCATATCATCTTGCAGAGAGAGATTTGAGTGGCTTGCTCAAC 1014

Db 1162 GCAAAATACAGATTTCCATATCATCTTGCAGAGAGAGATTTGAGTGGCTTGCTCAAC 1221

Qy 1015 AGGCTCTGGAAGAGATGCGGCTTTTCTCTACCAATTTTGGCTCATATGATGATGAG 1074

Db 1222 AGGCTCTGGAAGAGATGCGGCTTTTCTCTCTGCTCATATGATGATGAGATGAG 1281

Qy 1075 AATTAAGTCAAGTCTGTTTAAAGAGTGGAGAACCAAGATGATTAATTTGAGACCAAC 1134

Db 1282 AATACAGCGAGTGGCTTTAAAGAGCTGAGGAAACCAAGATGATATATGTAGCAAC 1341  
 Qy 1135 TCGAGAAATTTGTCACCAAGATTTATTTGGAAGCCAGAAAATTTCTTTGGGAGCTTGAT 1194  
 Db 1342 TCGAGAAATTTATCAACCAATTTATTTGGAAGCCAGAAAATTTCTTTGGGAGCTTGAT 1401  
 Qy 1195 AAGAGCTGTGTTATATATATGAGGAAACCCAGCTGGGACATTCATTGCACAAAATAGTACA 1254  
 Db 1402 AAGAGCTGTGTTATATATGAGGAAACCCAGCTGGGACATTCATTGCACAAAATAGTACA 1461  
 Qy 1255 AAGCTGTATATATATATATGAGGAAACCCAGCTGGGACATTCATTGCACAAAATAGTACA 1314  
 Db 1462 AAGCTGTATATATATATATGAGGAAACCCAGCTGGGACATTCATTGCACAAAATAGTACA 1521  
 Qy 1315 GATTGGTCTGAAACAGATCAAAATCTAGTTTGGATGAGTGAATGATGATGATAT 1374  
 Db 1522 GATTGGTCTGAAACAGATCAAAATCTAGTTTGGATGAGTGAATGATGATGATAT 1581  
 Qy 1375 GGGTTTGTGTCAGAAATGAAAGATTAATTTCTTGGCCAGAAATGCCATCAAGAAACA 1434  
 Db 1582 GGGTTTGTGTCAGAAATGAAAGATTAATTTCTTGGCCAGAAATGCCATCAAGAAACA 1641  
 Qy 1435 GCGCCAAACCTTATGTTCAAGTGAATTTCCAGAGAAATTCAAAAGTTGGCTGAGA 1494  
 Db 1642 GCGCCAAACCTTATGTTCAAGTGAATTTCCAGAGAAATTCAAAAGTTGGCTGAGA 1701  
 Qy 1495 GTTTTAAAGTCAATATCTGTTTGTGCTGTGCAAGAGTGGTGAACATGATAGA 1554  
 Db 1702 GTTTTAAAGTCAATATCTGTTTGTGCTGTGCAAGAGTGGTGAACATGATAGA 1761  
 Qy 1555 TGTTCAGAGACCTTCTCCAGTTGGCCAGTTCTCAAAAAGAAAGCTCGTTGAAT 1614  
 Db 1762 TGTTCAGAGACCTTCTCCAGTTGGCCAGTTCTCAAAAAGAAAGCTCGTTGAAT 1821  
 Qy 1515 TCTCGGAAACCTAGAGGAGTGAAGAACTATGTTGTTGAAATTAAGAAAAGAGA 1674  
 Db 1822 TCTCGGAAACCTAGAGGAGTGAAGAACTATGTTGTTGAAATTAAGAAAAGAGA 1881  
 Qy 1675 TTTTACTGCAACCTTTCTTTGTGCAAGAAAATATCAACTATCAAGTATTCAGCGATCG 1734  
 Db 1882 TTTTACTGCAACCTTTCTTTGTGCAAGAAAATATCAACTATCAAGTATTCAGCGATCG 1941  
 Qy 1735 GGAACAGAGAGAGGAGGAGCAAGCTTTGGAGATTTTGGCTTGGAAAGTGGCCAGTTCT 1794  
 Db 1942 GGAACAGAGAGAGGAGGAGCAAGCTTTGGAGATTTTGGCTTGGAAAGTGGCCAGTTCT 2001  
 Qy 1795 TGTGCTACTTCAGTAGCTGCGAGAGGCTTGATATTTGAAAATGTGCAACATGTTATCAA 1854  
 Db 2002 TGTGCTACTTCAGTAGCTGCGAGAGGCTTGATATTTGAAAATGTGCAACATGTTATCAA 2061  
 Qy 1855 TTTTGATTTCTCTTCTACCATTTGATGATATGTTTCAATGAAATTTGGGGTACTGCTGCTG 1914  
 Db 2062 TTTTGATTTCTCTTCTACCATTTGATGATATGTTTCAATGAAATTTGGGGTACTGCTGCTG 2121  
 Qy 1915 TGGGAATCTGCGAGAGCAATTTCTTTTGTGATCTTGAATCGATACGATTAACATTAGACA 1974  
 Db 2122 TGGGAATCTGCGAGAGGAGATTTCTTTTGTGATCTTGAATCGATACGATTAACATTAGACA 2181  
 Qy 1975 GCGCTTACTGAAAAGTATGACAGATGCTCAACAGAGATTTCTGATGATGTTTGAAGAAAT 2034  
 Db 2182 ACGCTTACTGAAAAGTATGACAGATGCTCAACAGAGATTTCTGATGATGTTTGAAGAAAT 2241  
 Qy 2035 TGGCTTTTGTGATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2091  
 Db 2242 TGGCTTTTGTGATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2301  
 Qy 2092 ATCAGTTGATACAGAA-----AAGGCAAGAGCACTTTGAACACAGCTGGGATTTTC 2142  
 Db 2302 ATCTTTTGACACTAGAGAAATTTCCAGGGCAAGAAACACATGAAACAGCTGGGATTTTC 2361  
 Qy 2143 TTCTTCAGAGCTCCCAATCCAGTAGATGATGATGATGATGATGATGATGATGATGATGATGAT 2200

Db 2362 TTTCGACAGCTCCCAATCCAGTTGATGACGAGTCAATGAGATTTAAAGCAACAGACATA 2421  
 Qy 2201 CTTCAAGTCTGTGATTT 2218  
 Db 2422 CTGCAAGCTGTATGTTT 2439  
 RESULT 8  
 MUSDVH  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE  
 AUTHORS  
 TITLE  
 Fujisawa, Y., Komiya, T., Kawabata, H., Sato, M., Fujimoto, H.,  
 Furusawa, M., and Noce, T.  
 Isolation of a DEAD-family protein gene that encodes a murine  
 homolog of Drosophila vasa and its specific expression in germ cell  
 lineage  
 Proc. Natl. Acad. Sci. U.S.A. 91 (25), 12258-12262 (1994)  
 MEDLINE  
 PUBMED  
 7991615  
 2 (bases 1 to 1930)  
 REFERENCE  
 AUTHORS  
 TITLE  
 Direct Submission  
 Submitted (01-Apr-1993) Yoshiaki Noce, Mitsubishi Kasei Institute  
 of Life Sciences, Developmental Biology, 11 Minamiooya, Machida,  
 Tokyo 194, Japan (Tel:81-427-24-6246, Fax:81-427-29-1252)  
 FEATUERS  
 source  
 1..1930  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="BALB/c"  
 /sub\_species="domesticus"  
 /db\_xref="taxon:10090"  
 /clone="mVH"  
 /cell\_type="primordial germ cell, spermatogonium,  
 spermatocyte"  
 /tissue\_type="gonad, testis"  
 /clone\_lib="lambda gt10"  
 /dev\_stage="adult, embryo"  
 <1..1914  
 /function="RNA helicase"  
 /codon\_start=1  
 /product="Drosophila vasa homologue"  
 /protein\_id="BA03584.1"  
 /db\_xref="GI:286075"  
 /translation="RGRGKGFGNRRGFLNNKFEKSPGSKWSENNDCEDNDRSGFS  
 KGGGCDNDNSASGPRRGRGGRGGRGGLRPNSESPODGTGCGGFTVLGX  
 PASDSGNDITQSRSGSGRGYKGLNEVVTGSGKMSKSTEGESDSGPGPVY  
 IPPPEDEDSFAHYQTGINFDKDTLLVEVSGDAPAILTFEBAKCYLLNNIR  
 KAGYTLFVQKXTTLPVLAGRDLMAACAQGTGKDAFLPLTLLHMMNDGTLASFKK  
 LOEPECILVAPRELINQYLEARKFSFCTCVISVYIGTGFHGSVROIVGCNLLC  
 ATPGRIMDIIGKXIKGVKLVLEADSDMSDMGFAPKIKLISCPGMSEKHQTL  
 LRSATPREIORLADPLKSNYLFVAVGVGAGADVDQTIIVGVGVKESKELRFEV  
 NIGDEKTMVFEVETKKADPIATFLCKEKSISISIGDPRRERDGLDFRGCKPVL  
 VASVAAKGLDIENVDVHNPDPSTIDIVHIGRTGRCGTATGALISFPDPSNNHL  
 APLVYVLSDAQDQVPAMLELIAFTYVPSFSSSTSGAVFASVDTKRYGKXKXHE  
 YSGDFEFTSSQSS"  
 ORIGIN  
 Query Match 64.9%; Score 1443; DB 10; Length 1930;  
 Best Local Similarity 86.4%; Pred. No. 0;  
 Matches 1671; Conservative 0; Mismatches 240; Indels 24; Gaps 6;  
 Qy 266 TTTCGAGTTGGAAGAATTTTGGAAACAGAGCTTTTCAACAGCAGGTTTGAAGATGGT 325  
 Db 1 TTTCGAAAGGAAAGGCTTTGGAACAGAGCTTTTAAATPACAAGTTTGAAGAAGGT 60

QY	1406	TCCTTCCCAAGAAATCCCATCCAAAGGAAACAGGGCCAAACCCCTTAAGTCAGACAACTTT	1466
Db	1135	TCCTTGTCCAGAAATCCCATCCAAAGGAAACAAACCCCTTTATTCAGTGCATCTTT	1199
QY	1466	CCAGAGAAATTCAAAGGTGTGGCTGCAGAGTCTTTTAAAGTCAAATATATCTGTGTGCT	1528
Db	1195	CCAGAAAGAAATCCAGAGGTGTGGCTGGGATTTCTTAAGTCCAAATATATTGTGTGTGT	1255
QY	1536	GTTGGACAAATGGGGTGGAGCATGTGAGAAATGTTCAAGCAGACCCGTTCTCCAAAGTTGGCCAG	1588
Db	1255	GTTGGCCAAATGGGAGGGAGCTTCGAGAAATGTTCAAGCAGAGAGATTCCTTCAAGTTGGCCAG	1314
QY	1586	TTCTCAAAAAAGAAAAGCTCGTTGAAATTTCTGGG-AAACATAGGGATGGAAGAACTAT	1644
Db	1315	-TATCAAAAAAGAAAAGCTCTGTGAGATTTCTAGAAAACATAGGTATGAAAAGACTAT	1372
QY	1645	GCTCTCTGTGAAACTTAAGAAAAAGCAGATTTACTGCAACTTTCTTTGTCCAGAAAA	1704
Db	1374	GGCTCTGTGTAAACCAAGAAAAAGCCGATTTCAATGCAACTTTCTTTGTCCAGAAAA	1433
QY	1705	AATATCAACTTCAAAGTATCCATGTGTATTCGGGAAACAGAGAGAGGGGAGACGCTCTGG	1764
Db	1434	AATATCAAGTACAAAGTATTCATGTGTATTCGGGAAACAGAGGAGAGAGCAGAGCTCTGG	1493
QY	1765	AGATTTTCGCTTTGGAAGAGTCCAGTTCTTGTGTACTTCACTAGCTGCACAGAGGCT	1824
Db	1494	AGATTTTCGCTGTGGAAGAGTCCAGTTCTTGTGTACTTCACTAGCTGCACAGAGGCT	1553
QY	1825	GGATATTTGAAAAATGTGCACATGTTATCATTTTGATCTCTCTTCAACATGTATGATA	1884
Db	1554	TGATATTTGAAAAATGTTCACATGTTATTCAAATTTGACCTCTCTCAACATGTATGATA	1613
QY	1885	TGTTCAATCGAATTTGGGGGTCTGGTGCCTGTGGGAAATACGGGACAGCAATTTCTCTTT	1944
Db	1614	TGTTCAATCGAATTTGGAGCGCACTGGCCGCTGTGGGAAATACGGGAGAGGATTTCTTTT	1673
QY	1945	TGATCTTGAATTCGANTAACATTATAGACAGCCTCTATGATAAAGTATTTGACAGATGCTCA	2004
Db	1674	TGATACAGACTCTGATATATCATTTAGACAGCCCTCTATGATAAAGTATCTGACAGCGCTCA	1733
QY	2005	ACAGAGATGTTCTTGCAATGTTTGGAGAAATTCGCTTTAGTACATACAT--TCTGTGCTT	2061
Db	1734	ACAGAGATGTTCTCCCGCACTGGCTAGAGAGATTTGCTTAGTACATACATATGCTCCCACTT	1793
QY	2062	CAGTGTAGTATCAAGA---GAAAACGTTGTGCATCGTGTGATCCAGAAA-----	2109
Db	1794	CAGTGTAGTATCAAGAAGGGGGGGCCGTGTTTGATCTGTTGACAGAGAGAAATTTACCA	1853
QY	2110	-GGGCAAGAACACTTTTGAACACACAGCTGGGTTTTCTTTACAGAGCTCCCAATCCAGTAG	2168
Db	1854	GGGCAAGAACACAGTTGAATTCAGCCGGGATTTCTTTCACAAAGCTCCCAATCAATTTG	1913
QY	2169	ATGATGTAGTATGGG 2183	
Db	1914	ATGACGATATCATGGG 1928	

RESULT 9	AE004836	1989 bp	mRNA	linear	VRT 01-SEP-2000
LOCUS	AB004836				
DEFINITION	Gallus gallus mRNA for Cvh, complete cds.				
ACCESSION	AB004836				
VERSION	AB004836.1	GI:9967267			
KEYWORDS	Cvh.				
SOURCE	Gallus gallus (chicken)				
ORGANISM	Gallus gallus				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.				
REFERENCE	1 (sites)				
AUTHORS	Tsunekawa, N., Naito, M., Sakai, Y., Nishida, T. and Noze, T.				
TITLE	Isolation of chicken vasa homolog gene and tracing the origin of primordial germ cells				

JOURNAL Development 127 (12), 2741-2750 (2000)  
 MEDLINE 20283536  
 PUBMED 10821771  
 2 (sites)  
 Tsunekawa, N., Fujimoto, H., Nishida, T. and Noce, T.  
 Isolation of a DEAD-family protein gene that encodes a chicken  
 homolog of Drosophila vasa and its specific expression in germ cell  
 lineage  
 JOURNAL Unpublished  
 REFERENCE 3 (bases 1 to 1989)  
 AUTHORS Tsunekawa, N.  
 TITLE Direct Submission  
 JOURNAL Submitted (13-JUN-1997) Naki Tsunekawa, Mitsuishi Kaei Institute  
 of Life Sciences, Molecular Reproduction and Development; 11  
 Minamiooya, Machida, Tokyo 194, Japan  
 [E-mail: tsunek@libra.1s.m.kagaku.co.jp, Tel: +81-427-24-6246,  
 Fax: +81-427-24-6316]  
 Location/Qualifiers  
 1. 1989  
 /organism="Gallus gallus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9031"  
 1. 1989  
 /codon\_start=1  
 /product="Cvtn"  
 /protein\_id="BAB12337.1"  
 /db\_xref="GI:9967268"  
 /translation="MEEDMDTELEQENAAAGSSEQAMMANSGRNPSPLRFSRSP  
 SPSLPSPGRPNPSFSGSONKSLANEINLSLPVQHDIGYSGSRSVRNQED  
 OPIYRGRGRSGSRDROENRNSANDPQMDQGRVAVGIGOSKCNSENRNPLGS  
 PPAQGRGAVGSPAGVTKRSEIDSRGKTYVPPRPEDRPSI.PACVQGINDK  
 YDEAVMSGLDPPAPLAFEEANFATLKNISKTYEKLTVQGHSLPVIDAGDL  
 MSAQTSQSGTAFLPIVDKMKDVTAFSPQDPOCIIVAPTELNQIFLEKXK  
 FVYGTICRPVYIGVQTHSIRIQMGSCNLCATPGRFLDIIEKRSISLVEKYL  
 DEADRMIDMGFGLDMKLI.SYPMPSKDRQITMPSATFEEVORLAGEFLKDYFL  
 VIENTGAGSDVQONILQVPRLSKRDLLIILQSTGERMTVPVDRKKADYLAFLC  
 QENLPSTSHGDRQPRRRLALDRPSGKQILIVANSVSRGLDIENVQVINFDPN  
 TIEDYVRIQRTGRGNTGRKAVSEFPDQSGHLVQSLIKVLSLTQOEPRQSGMAYQR  
 TNIVASTWCPRLQAVAVENPFRMRKSYSETTFKWE"

ORIGIN  
 Query Match 33.8%; Score 751.2; DB 5; Length 1989;  
 Best Local Similarity 71.1%; Pred. No. 1,7e-167;  
 Matches 1009; Conservative 0; Mismatches 408; Indels 3; Gaps 1;

Db 661 AGAAGTAATAACAGGCTCTGGAAAGAAATCTTGAAGTCAGAACGACAGAGAGAGAAAG 720  
 492 AGGAGAAAGAGAGACAGTTGGAGGTCTCTGCAGAGATTCTCAAGAGAGCTCTGAAGAAAT 551  
 721 TAGGATATCTCAAGAGACCAAGAGTACCTACATACCCCTCTCCACCTGAGATGAGA 780  
 552 TGAATTCGGAAGAGGTCCAAAGGTGACTTATGTCCCCCTCTCTCCACTGAAGTGAACA 611  
 781 CTCATCTTTGCACATTAACAGACGACATTAACCTTGACAAATACGACATATCTTGT 840  
 612 GTCCATCTTTGCATGATACAGTACAGGAAATTAATTTGACAGATGATGATGTGCTGT 671  
 841 GGAAGTCTTGACATGATGACACACGACGCAATTCGATCTTTGAAGAAGTATATCTCTG 900  
 672 TGAATGTCAGAGACTTGACCTCCAGACACCAATTAAGCTTTTAAAGAGTAACTTTGC 731  
 901 TCAGACATGATTAACAACATTTGCTAAAGCTGTTACTAAGCTTACTCTCTGTGCAAAA 960  
 732 TCAGCTTTAAGAAAGATATATCTAAACCTGATATCAAAAATCTTACTCAGTGAGAA 791  
 961 AATACGATATCTCTATCACTTGAAGACGAGATTTGATGCTGTGTCTCAAAAGGCTC 1020  
 792 GCACAGCATTCCTGTTATACAGAGCGGGGATTATATGTCATGTCGCCAGACAGATC 851  
 1021 TGGGAAGACTGGGCTTTTCTCTCACAATTTGGCTCATATGATGATGATGATGATGATGAT 1080  
 852 AGGAAGAAACAGAGCTTTTCTTCTACCAATTTGTGACCGGATGATGATGATGATGATGATGAT 911

1081 TGCCAGTCGTTTAAAGATTGACAGAAACAGAGTATATTGTGACACCAACTCGAGA 1140  
 912 TGCAAGC---TTCCAAAGCAGAACCCCAATGATGATATTGTGACCACTGAGAGA 968  
 1141 ATGTGTCAACAGATTTATTTGGAAGCAGAAAAATTTCTTTGGGACTTGTGTAAGGC 1200  
 969 ACTGATTAATCAAGATCTTCTTAAGACAGAAAGTTGTGTATGGGACTGTATAGGCC 1028  
 1201 TGTGTATATATAGGGGAAACCCAGCTGGGACATTCATTTGACAAATATGACAAAGCTG 1260  
 1029 TGTGTATATATAGGAGTACACAGACAGTCAATTCATTCGTCATTAATGCAAGCTG 1088  
 1261 TAATATATATATGCTACTCTCTGGAAGCTGATGATATCATATAGCAAAAGAAAGATTGG 1320  
 1089 TAATATATATATGCTACTCTCTGGAAGCTTCTTGAACATTAATGAAAAAGGAAAGATAG 1148  
 1321 TCTCAACAGATCAAAATCTAGTTTGGATGAGCTGATGCAATGTTGATATGGTTT 1380  
 1149 TTTGGTGAAGGTGAATAATTTGCTAGTATGAGACAGACCGCATGCTCATATGGTTT 1208  
 1381 TGGTCCAGAAATGAAGAAAGTTAATTTCTTGCCCGAAGAAATGCCATCAAAAGACGCCCA 1440  
 1209 TGATTTAATATGAAGAAAGCTGATTTCTTATCCAGAAATGCCATCTAAAGACAGCTCA 1268  
 1441 AACCTTATGTTCAAGTGAACATTTCCAGAGAAATTAAGGTTGGCTGACAGATTTT 1500  
 1269 AACATTAATGTTAGTGAACATTTCTCTGAGAGAAATTAAGGCTGCTGATGATTTT 1328  
 1501 AAAGTCAATATATCTGTTTGTGCTGTTGGAACAATGGGTGAGAGATGATGATTTCA 1560  
 1329 GAAAACGAGCTATATATTTCTGTTATGGAATACCTGTGAGAGCTGACAGATGTTCA 1388  
 1561 GCAGACCGTCTCCAGATTTGCGCAATTTCCAAAAGAAAGCTCGTTGAATTTCTGCG 1620  
 1389 GCAAAATATCTTCAGGTTCCCGGATATCCAAAGGAGATTAATTAAGAAATTTCTCA 1448  
 1621 AAACATAGGGAGTAAAGAAAGTATGATGCTTGTGTAACCTAAGAAAAAGCAGATTTTAC 1680  
 1449 AAGCAGAGTGTGACAGAACCATGATGTTGTGACCAAAAGAAAAAGCAGATTTACCT 1508  
 1681 TGCAACTTTCTTTGTCAAGAAAAATATCAACTTCAAGTATCATGATGATCGGGAACA 1740  
 1509 TCAGAGCTTTCTTTGTCAAGAAACCTTACCTACCCAGCAATTCATGAGATGAGGAACA 1568  
 1741 GAGAGCGGGAGCAAGCTCTGAGATTTTGGCTTTGGAAGTCCCAAGTCTTGTGC 1800  
 1569 GAGAGAGAGAGATAGCTCTTCCGATTTCCGTTCTGGAATATGCAATTTCTGTGCGC 1628  
 1801 TACTTCAATAGCTGCAGAGAGGCTGATATTAAGAAATGTCACATGTTATCAATTTTGA 1860  
 1629 AACTTGGATGATCAAGAGGCTGATATTAAGAAATGTTCAACATGTTATTAATTTTGA 1688  
 1861 TCTTCTTTTACATTTGAGAAATATGTTATGATGAAATTTGGGCTTACCTGCTGTGGGA 1920  
 1689 TCTCTTCAACCATTAAGATTAATGATGATGAAATTTGACCACTGCTGTGGGA 1748  
 1921 TACTGCAAGCAATTTCTTTTGTGATCTGATGATGATGATGATGATGATGATGATGATGAT 1800  
 1749 TACTGCAAGCAAGATTTTCAATTTGATGATGATGATGATGATGATGATGATGATGATGAT 1808  
 1981 AGTAAAGATTTGACAGATGCTACAGAGATGTTCTGATAGTGTGGAAGAAATTTGCTT 2040  
 1809 ACTTAAAGTCTTTCCAGAACCCAGAGGAATTTCCAGTTTGTGGAAGATGAGCTGTCCA 1868  
 2041 TAGTACATCATCTCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2080  
 1869 AGAACAATATTTGTTGCTTCAACTGTGTGCCAAAGGGA 1908

RESULT 10  
 AF046043 2502 bp mRNA linear VRT 19-FEB-1998  
 LOCUS AF046043  
 DEFINITION Xenopus laevis DEAD box protein (VLG1) mRNA, complete cds.







Qy 1855 TTTTGATCTTCTTCTTACCATTTGTAATGTTTCATGCAATTGGCGGTAAGTGGCTTG 1914  
 Db 1879 TTTTGACATGCGCCAGCGACATGATGATGTCATCGATAGAGAACTGGAAGCTG 1938  
 Qy 1915 TGGGAATACCTGGCAGAGCAATTTCTTTTGTGATCTTGAATCGGATTAACATTTCACAA 1974  
 Db 1939 TGGGAACACCGGTGGCGCGCTGTCTTTTAAACCCAGATCTGACATCTCATCTG 1998  
 Qy 1975 GCCTCTAGTAAAGATTTGACAGATGCTCAACAGAGATTTCTGCAATGTTGGAGAAAT 2034  
 Db 1999 CTCCTGTGTAAGTCTTTTCAGGGGCCCAAGAGTTCGAAATGCTGGAGAGT 2058  
 Qy 2035 TGCCTTAGTACATACATCTCTGCTCAGTGTGATGACAGAGGAAAGCTGTTTCATC 2094  
 Db 2059 TGCTTCAAGTCTCA---TGACACACAGGCTTCAACCCAGTGGAAGGTTTGCATC 2115  
 Qy 2095 AGTTGATACCAAGAGG 2112  
 Db 2116 TACAGACTCGCGAGAGG 2133

RESULT 12  
 LOCUS DRY12007 2492 bp mRNA linear VRT 09-OCT-1997  
 DEFINITION D. rerio vlg (vasa like gene) mRNA for putative RNA helicase.  
 ACCESSION Y12007  
 VERSION Y12007.1 GI:2558534  
 KEYWORDS RNA helicase; vasa gene; vlg gene.  
 SOURCE Danio rerio (zebrafish)  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.

REFERENCE  
 AUTHORS Olsen, L.C., Aasland, R. and Fjose, A.  
 TITLE A vasa-like gene in zebrafish identifies putative primordial germ cells  
 JOURNAL Mech. Dev. 66 (1-2), 95-105 (1997)  
 MEDLINE 98025484  
 PUBMED 9376327  
 REFERENCE 2 (bases 1 to 2492)  
 AUTHORS Olsen, L.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (24-MAR-1997) L.C. Olsen, University of Bergen, Department of Molecular Biology, Thormoehensgt 55, N-5020 Bergen, NORWAY

FEATURES  
 source  
 1..2492  
 /location/Qualifiers  
 /organism="Danio rerio"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:7955"  
 /dev\_stage="embryonic, late blastula"  
 1..2492  
 /gene="vlg"  
 35..2138  
 /gene="vlg"  
 /note="vasa-like gene putative"  
 /codon\_start=1  
 /evidence=not experimental  
 /product="RNA helicase (DEAD box)"  
 /protein\_id="CAA72735.1"  
 /db\_xref="GI:2558535"  
 /db\_xref="GOA:O42378"  
 /db\_xref="SPTREMBL:O42378"  
 /translation="MDWEDQSPVAVSGSGFGAGANDKSNSEGTGSSWMTGDSFR GRGGRGSGRGGFGSGFKSIEDGNSDGNNGSGRGRGFRGSGRSGRDNDE NNDNDWKSGRGRGGRGGFGGFGPDGDNEDTGRGGRGNENNGENDEGGEG RGRGGRGFRGGGDESGRGRGGRGRGRNEVFSYTTADKIDRSGSNAPKIV YVPPPPREPSIPSHYATGTFDKYDILVDVSGSNPKAIMTFEAGLSDLSKIV SSGYVKTPTVQKGIPLTISGRDLMACQSGSKTAAFLPLPIQRFETDVAASKS EIOEPBIIIVATRELINQIYLEARKFAYGTCVPRVVYGGINTGYITREVLKGCNVL

## protein\_bind

Query Match 30.9%; Score 686.4; DB 5; Length 2492;  
 Best Local Similarity 69.6%; Pred. No. 4.2e-152;  
 Matches 927; Conservative 0; Mismatches 405; Indels 0; Gaps 0;  
 /note="similar to nanos response element"  
 /bound\_moiety="a putative nanos-dependent factor"  
 /evidence=not experimental

## ORIGIN

CATPGRUHLILGRKIGISIKRYVYLDEADRMIDMFEPEMERKIVASPGMSKEKROT  
 LMFATYPEDIQRNAADFLKVDYIFLAVGVGACSDVEQITVQVDYSKRDQLELL  
 RATGERMVFVETKRSADFLATFLCOEKISTISIHGDREREKALSPRLGHCYV  
 LVATSVARGDIDVOYVNFDMPSIDIEVNHIGTGCGNTRGAVSPFNPSDPR  
 LARSLIKVLSGAQOVPRKMLEVAFSHGTTGNPRKVAISTDSRGSGFKSDPEPP  
 SOTAPSAALADDEWE"  
 2255..2277  
 /gene="vlg"  
 /note="similar to nanos response element"  
 /bound\_moiety="a putative nanos-dependent factor"  
 /evidence=not experimental

Qy 715 AGAAGTATGATCTCAAGAGCAAGCAAGTACATATCCCTCTCCACCTGAGGA 774  
 Db 656 AGAAGAGTGAATGCAAGCAAGCAAGTGTGTTATGTCGCGCGCTCTCCAGAGA 715  
 Qy 775 TGAGACTCCATCTTTGCACTTATCAGACGCGATPAACTTGCAAAATACACTAT 834  
 Db 716 GGAAGCTCGATATCTCCCATTAATGCAAGCGCATTAATTTGACAAATATGATAT 775  
 Qy 835 TCTTGGAAGTGTCTGGAACATGATGACACCAACGCAATTCGACTTTGAGAAGCTAA 894  
 Db 776 TCTTGGAAGTGTCTGGAACATGATGACACCAACGCAATTCGACTTTGAGAAGCTAA 835  
 Qy 895 TCTGTGACAGACTGATTAACAACATGTGTAAGCTGTATATCAAGCTTACTCTGT 954  
 Db 836 ACTTGTGACACTGATGACAAATGATCAAGCTGTGATATGTAAGCTTACTCTGT 895  
 Qy 955 GCAAAATATACATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1014  
 Db 896 TCAAAACATGCACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 955  
 Qy 1015 AGGCTCTGGAAGACTGCGCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1074  
 Db 956 TGATCAAGAAAGAGCGCGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1015  
 Qy 1075 AATAACTGCCAGTCTTTTAAAGTTCGAGAACCAAGATGATATTTATGACCAAC 1134  
 Db 1016 TGTCGAGCCAGCAAGTTCAGTATGATGAGAGGAGCCGATATCTGCTCCAC 1075  
 Qy 1135 TCGAATATGTCACAGATTTATTTGAAGCCAGAAATTTCTTTTGGAGCTGTGT 1194  
 Db 1076 CAGAGACTTATCATATGATCTATCTGAGGCCAGAAATTTGATACGAGACTGTGT 1135  
 Qy 1195 AAGACTGTTGTTATATATGAGGAGAACCGCTGGGACATTCAATTGACAAATAGTACA 1254  
 Db 1136 ACGACCTGTGTGTTATGAGAGTATTAATACCGATATATCTATCCGAGAGCTGTAA 1195  
 Qy 1255 AGGCTAT 1314  
 Db 1196 GGGCTGATGAT 1255  
 Qy 1315 GATTGCTCTGCAAGATGAGTATATATATATATATATATATATATATATATATATATAT 1374  
 Db 1256 GATTGCTCTGCAAGATGAGTATATATATATATATATATATATATATATATATATATAT 1315  
 Qy 1375 GGGTTTGGTCCAGAAATGAAAGTAAATTTTTCGCCAGAAATGCCATCAAGAGAA 1434  
 Db 1316 GGGTTTGGTCCAGAAATGCGCAACTGGTGGGCTCTCTGATATCTTCAAAAGAGA 1375  
 Qy 1435 GCGCAAAACCTTATGTTCAAGTCACTTTTCAGAGAAATTCAAAGTGGCTGACAG 1494  
 Db 1376 ACGCAAAACCTTATGTTCAAGTCACTTTTCAGAGAAATTCAAAGTGGCTGACAG 1435  
 Qy 1495 GTTTTAAAGTCAAT 1554  
 Db 1436 CTTTCTTAAGTCAAT 1495









755 CGGATCCCATCATTTCTGCTGAGAGGATCTCATGCTGCGCCCAAGCTGGCTGG 814  
1024 GAAGATGCGGCTTTTCTCCCTACCAATTTTGGCTCATATGATGATGAAATATCTGC 1083  
815 AAAAGGGCTGCTTCTGCTCCCAATCTTCAACAGCTGATGGCAGATGGTCGGCGGC 874  
1084 CAGTCGTTTAAAGAGTTCAGAACAGAGTGTATTTTGTAGCAGCACTCGAATTT 1143  
875 CAGTTCCTGAGAGATGAGAGAGCTGATGCTCATATTTGTGGCAACACTAGGGAGCT 934  
1144 GGTCAACAGATTTTATTTTGGAGCGAAGAAATTTCTTTTGGACTTGTATAGAGCTGT 1203  
935 GATCAACAGATTTTACCTGAGGCGAGAGAAATTTTCCATGAGCAATGTGTCCACTGT 994  
1204 TGTATATATATGAGGAGAACCCAGCTGGGACATTCATTCGCAATAGTCAAGCTGTAA 1263  
985 TGTATGTATATGAGAGTACAGCACTGGACACCAATTAAGAGCTCTTAGGGGCTGTAA 1054  
1264 TATATATGCTGCTACTCTGAGAACTGATGATATCTATAGCAAGAAAGATGCTCT 1323  
1055 TGTGTGTGTGAGAACACAGAGCACTGCTGACATGATGAGAGAGAAAGTGGCTT 1114  
1324 CAAACATCAAAATCTTATGATTTGATGAGTACATTCGCAATGATGATGGTTTGG 1383  
1115 GACTAAGTCCGTTACTGTGCTGATGAGGCTGACGCAATGTGATATGGGTTTGA 1174  
1384 TCGAAGATGAGAAATTAATTTCTGCTCCAGGATGCCATCAAGGAAACGCGCAAC 1443  
1175 GCTGACATGCGCGCGCTGGTGGCTCTCTGAGGATCCAAAGAGAACCGTCAGAC 1234  
1444 CTTATGTTCACTGACATTTTCCAGAGAAATTCAAAGTGGCTGACAGATTTTAA 1503  
1235 TGTATGTATGATGATGATTTCCCAAGGACATCCAGAGCTTGCAGCTGACTTTAA 1294  
1504 GTCAAAATTAATCTGTTTGTGCTGTGGACAGTGGGTGGAGCATGTAGATGTTCAGCA 1563  
1295 GACTGACTATTTGTTCTTGGCGGATTTGTGGGTGGAGCTTGCAGTGTATGGAGCA 1354  
1564 GACCGTTCTCAAGTTGCGCCAGTTCTCAAAAAGAGAAAGCTGTTGAATTTCTGCCAA 1623  
1355 GACATTTGTGAGGTACACCAAGTTCCAAAAGAGAGAGAGCTATATGACCTCCGAATTC 1414  
1624 CATAGGAGATGAGAACTATGTTCTTTTGTGAATCTAAGAAAAAGAGATTTTACTGC 1683  
1415 AACAGATCGAGCGCACATGGTGTGTAGAGCCAGAGACAGGCTGATTTTATTTGC 1474  
1684 AACTTTTCTTTGTCAAGAAAAATATCAACTACAGATATCCATGTGATTCGAGAACAGAG 1743  
1475 CACATCTTGTGTGAGAGAAATTTCCAACTACAGCATTCACGGTGAACCGTGAAGCTG 1534  
1744 AGAGCGGAGCAAGCTCTTGGAGATTTTGGCTTGGAAAGTGGCCAGTTCTTGTGCTAC 1803  
1535 GAGCGAGAGAGAGGCGCTTGGAGATTCGCTGGCAATGTTCTGTCTGGTGCAC 1594  
1804 TTCAATAGCTGCGAGAGGCTGAGATTTGAATGTGCAACATGTTATCAATTTGATCT 1863  
1595 CTCTGTAGGTGCCCGGAGCTGATATTCAGATGTGCAATGTGTGTGAACTTTGACT 1654  
1864 TCCCTTACCATGTAATGATATGTTGATGATGAGGCTGATGCTGTGTGGGAATAC 1923  
1655 CCCAACAATATCGACGAATGTGTCATGTATGTGTAACCGGCGGCTGTGGCAAC 1714  
1924 TGGCAGAGCAATTTCTTTTGTATCTTGAATCGATACCATTTAGCAGAGCTCTAGT 1983  
1715 CGGAGAGGCAATGTCTTCTATATATCAGAGAGCTGACAGTGAATGGCTCATCTGGT 1774  
1984 AAAAGTATGACAGTGTCTACAGAGATTTCTGCAATGGTGTGAAAGAAATTTGCTTTAG 2043



JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9585-9590 (2000)  
 MEDLINE 20402578  
 PUBMED 10920202  
 REFERENCE 2 (bases 1 to 2224)  
 AUTHORS Castillon, D.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (04-JUL-2000) Pathology, Brigham and Women's Hospital, 75 Francis Street, Boston, MA 02115, USA  
 FEATURES  
 SOURCE 1. 2224  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /chromosome="5"  
 /map="5q11.2-12"  
 /tissue\_type="testis"  
 /dev\_stage="adult"  
 1. 2224  
 /gene="VASA"  
 14. 2188  
 /gene="VASA"  
 /product="DEAD box RNA helicase"  
 /protein\_id="AF86585.1"  
 /db\_xref="GI:9438227"  
 /translation="MGEDMEAEINPHMSYVPIPEKRYSGENGDNENRTPASSSEM  
 DPGSPRDHMKSGPAGFCGRDAGCNKRDNTSGRGVGSFGNRGSRFE  
 DGSSGFWBSSSDCDNPNRNGFSKRGYDNNSESGYRRGSGSFRGCRGF  
 GLGSNNDDLPDRCQRTGLFGRSRPVLSGTGNDTSGRSSESGYKGLNEV  
 ITSGKNMSSENGESSDTQPKVYLPPEPDEDEIFAHYGTGINFDKXDTILV  
 EVSGHDAIPALITFEENALCOLNNNIAAGYKLPVQKYSIPILRGDLMAQOT  
 GSGTAAFLPLIAHMHGDIPTASRKELOEPECIIVATRELNVQIYEAKFSFGT  
 CVRAVYGGTOLGHSIROIVGSCNLTCPGRIMDIKEKIGIKOIKYVLEADNR  
 MLDMGFGEPMKILISCPMSEKROBTLMFSAPEPEIORLAPELKUNKYLVAAGD  
 GACRDVQVTVLOVGFSPKREKVELLRNIGDRTMVPFTKKADPFRTPFCQKIS  
 TSIHDDRERERQALGDFRFGKCPVATVATVAAGDLINVOHINIDPSTIDEX  
 VRIIGTRGNGTGRAISPFDELSDNHLAQPLVKVLTDAQDQVPMLEIARSTYIPG  
 FSSSTRGNVFAVDYTRKKSITLTAAGFSSSRANPNVDESDM"

## ORIGIN

Query Match 100.0%; Score 2224; DB 9; Length 2224;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2224; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACTTGAAGTACATGGGGGATGAAGATTGGAGCAAGAAATCAACCTCATATGCTTC 60  
 Db 1 ACTTGAAGTACATGGGGGATGAAGATTGGAGCAAGAAATCAACCTCATATGCTTC 60  
 QY 61 CTATGTTCCCATATTGAGAGAGATAGCTATTCTGAGAAAATGAGACAAATTTTACAG 120  
 Db 61 CTATGTTCCCATATTGAGAGAGATAGCTATTCTGAGAAAATGAGACAAATTTTACAG 120  
 QY 121 GATCCAGCTTCATCATCAAGAAATGATGAGACCTTCGAGAGAGATCATTTCTAGA 180  
 Db 121 GATCCAGCTTCATCATCAAGAAATGATGAGACCTTCGAGAGAGATCATTTCTAGA 180  
 QY 181 AAGTGATTTGCTCTGGGCGGAATTTTGGAAAAGAGATGCTGTGAGTGTATTAAGCG 240  
 Db 181 AAGTGATTTGCTCTGGGCGGAATTTTGGAAAAGAGATGCTGTGAGTGTATTAAGCG 240  
 QY 241 AGATTAATACATCAACATGGGTGTTTGGAGTTGAGAAAGATTGGAAAAGAGGCTT 300  
 Db 241 AGATTAATACATCAACATGGGTGTTTGGAGTTGAGAAAGATTGGAAAAGAGGCTT 300  
 QY 301 TTCAACAGCAGGTTTGAAGATGATAGCTCTGCTTCTGAGAGAGCTTATGATGA 360  
 Db 301 TTCAACAGCAGGTTTGAAGATGATAGCTCTGCTTCTGAGAGAGCTTATGATGA 360  
 QY 361 CTGCGAAGATTAATCCACAGGAGAGAGGTTTCCACAGAGGCGGCTTATCGAGATGG 420  
 Db 361 CTGCGAAGATTAATCCACAGGAGAGAGGTTTCCACAGAGGCGGCTTATCGAGATGG 420  
 QY 421 AAATTAATTCAGAGCTTCAGGCGCATACAGAAAGAGTGAAGAGTATGTTCCAGAGTTG 480

Db 421 AAATTAATTCAGAGCTTCAGGCGCATACAGAAAGAGTGAAGAGTATGTTCCAGAGTTG 480  
 QY 481 CCGTGAAGATTTGCTTAGAAGTCCAAATTAATGATAGACCCAGACCAATGATATGA 540  
 Db 481 CCGTGAAGATTTGCTTAGAAGTCCAAATTAATGATAGACCCAGACCAATGATATGA 540  
 QY 541 GCGCACTGGGCGCTTTTGGCTTTTAAAGACCAAGTTTAAAGTGCACAGTATGTTGA 600  
 Db 541 GCGCACTGGGCGCTTTTGGCTTTTAAAGACCAAGTTTAAAGTGCACAGTATGTTGA 600  
 QY 601 TACTTCTCAAGCAGAGTGCAGTGAAGTGAACGAGTGTATCAAAAGTTTAAATGA 660  
 Db 601 TACTTCTCAAGCAGAGTGCAGTGAAGTGAACGAGTGTATCAAAAGTTTAAATGA 660  
 QY 661 AGAAGTAAATAACAGCTCTGAAAAGAAATTTTGGAGTCAAGACGAGAGAGAGAG 720  
 Db 661 AGAAGTAAATAACAGCTCTGAAAAGAAATTTTGGAGTCAAGACGAGAGAGAGAG 720  
 QY 721 TAGGATATCTCAAGACCAAAAGTACATACCCCTCTCCACCTGAGAGTGAAGA 780  
 Db 721 TAGGATATCTCAAGACCAAAAGTACATACCCCTCTCCACCTGAGAGTGAAGA 780  
 QY 781 CTCCATCTTTGACATTATCAGACAGGCAATAACTTGACAAATACGACATATCTTGT 840  
 Db 781 CTCCATCTTTGACATTATCAGACAGGCAATAACTTGACAAATACGACATATCTTGT 840  
 QY 841 GGAAGTCTGAGAAATATGACACACAGCAATTCGACCTTTGAAAGAACTATCTCTG 900  
 Db 841 GGAAGTCTGAGAAATATGACACACAGCAATTCGACCTTTGAAAGAACTATCTCTG 900  
 QY 901 TCAGACACTGATTAACAACATTTGCTAAAGCTGTATTAAGCTTAAGCTTAAGCTTA 960  
 Db 901 TCAGACACTGATTAACAACATTTGCTAAAGCTGTATTAAGCTTAAGCTTAAGCTTA 960  
 QY 961 ATACAGATTTCTATCTACTTGAAGACAGATTTGATGCTGTGCTTAAACAGGCTC 1020  
 Db 961 ATACAGATTTCTATCTACTTGAAGACAGATTTGATGCTGTGCTTAAACAGGCTC 1020  
 QY 1021 TGGAGACTGCGGCTTTTCTCCATCAATTTTGGCTCATATGATGATGATGAATTAAC 1080  
 Db 1021 TGGAGACTGCGGCTTTTCTCCATCAATTTTGGCTCATATGATGATGATGAATTAAC 1080  
 QY 1081 TGCCAGTCTGTTTAAAGATGAGAGACAGAGTGTATTTGATGATGATGATGATGAT 1140  
 Db 1081 TGCCAGTCTGTTTAAAGATGAGAGACAGAGTGTATTTGATGATGATGATGATGAT 1140  
 QY 1141 ATTGCTAACAGATTTATTTGAGAGCCAGAAATTTCTTTTGGACTGTGTAGAGC 1200  
 Db 1141 ATTGCTAACAGATTTATTTGAGAGCCAGAAATTTCTTTTGGACTGTGTAGAGC 1200  
 QY 1201 TGTGTTATATATGAGGGAACCCAGCTGGGACATTCATTCGACAAATAGTACAGGCTG 1260  
 Db 1201 TGTGTTATATATGAGGGAACCCAGCTGGGACATTCATTCGACAAATAGTACAGGCTG 1260  
 QY 1261 TAAATATATATGCTACTCTCGGAGACATGATGAGATATCATAGGCAAAAGAAATGG 1320  
 Db 1261 TAAATATATATGCTACTCTCGGAGACATGATGAGATATCATAGGCAAAAGAAATGG 1320  
 QY 1321 TCTCAACAGATCAAAATAGTATGTTGATGAGAGTATGATGATGATGATGATGATG 1380  
 Db 1321 TCTCAACAGATCAAAATAGTATGTTGATGAGAGTATGATGATGATGATGATGATG 1380  
 QY 1381 TGTGTCAGAAATGAAGAGTTAATTTTGTGCCAGAGATGCCATCAAGAGACAGCGCCA 1440  
 Db 1381 TGTGTCAGAAATGAAGAGTTAATTTTGTGCCAGAGATGCCATCAAGAGACAGCGCCA 1440  
 QY 1441 AACCTTAATGTTTCAAGCACTTTTCCAGAGAAATCAAGAGTGGCTCAGAGTTT 1500  
 Db 1441 AACCTTAATGTTTCAAGCACTTTTCCAGAGAAATCAAGAGTGGCTCAGAGTTT 1500  
 QY 1501 AAAGTCAATTAATCTGTTGTTGCTGTGAGCAAGTGGTGAAGCATGTAAGATGTTCA 1560

Db 1501 AAAGCAATTATCTGTTGTTGTTGCTGTTGGAAGAGTGGGTGAGCAGTATGATGTTCA 1560  
QY 1561 GCAGACCGTTTCCAGAGTTGGCCAGTTCTCAAAAAGAGAAAGCTGCTGAAATTTGCG 1620  
Db 1561 GCAGACCGTTTCCAGAGTTGGCCAGTTCTCAAAAAGAGAAAGCTGCTGAAATTTGCG 1620  
QY 1621 AAACATAGGGATGAAAGAACTATGCTTTTGTGAAATCTAGAAAAGAGCAATTTTAC 1680  
Db 1621 AAACATAGGGATGAAAGAACTATGCTTTTGTGAAATCTAGAAAAGAGCAATTTTAC 1680  
QY 1681 TGCAACTTTTCTTGTCTCAAGAAAAATATCACTACAAAGTATTCATGGTGTATGGGAGCA 1740  
Db 1681 TGCAACTTTTCTTGTCTCAAGAAAAATATCACTACAAAGTATTCATGGTGTATGGGAGCA 1740  
QY 1741 GAGAGAGCGGAGCAAGCTCTTGGAGATTTTGGCTTTGAGAAAGTGGCCAGTTCTTGTTC 1800  
Db 1741 GAGAGAGCGGAGCAAGCTCTTGGAGATTTTGGCTTTGAGAAAGTGGCCAGTTCTTGTTC 1800  
QY 1801 TACTTCAGTACTGCTCCAGAGGCTGGATATTTGAATTTGCAACATGTTATCAATTTTGA 1860  
Db 1801 TACTTCAGTACTGCTCCAGAGGCTGGATATTTGAATTTGCAACATGTTATCAATTTTGA 1860  
QY 1861 TCTTCTCTTCTACCATGATGATATGTTCAATGCAATTTGGGCGTACTGCTGTTGGGAA 1920  
Db 1861 TCTTCTCTTCTACCATGATGATATGTTCAATGCAATTTGGGCGTACTGCTGTTGGGAA 1920  
QY 1921 TACTGGAGAGCAATTTCTTTTGTGATCTTGAATCGATTAACCATTTAGACAGCTCT 1980  
Db 1921 TACTGGAGAGCAATTTCTTTTGTGATCTTGAATCGATTAACCATTTAGACAGCTCT 1980  
QY 1981 AGTAAAGTATTTGACAGATGCTCAAGAGATGTTCTGCAATGTTGGAGAAATTTGCTT 2040  
Db 1981 AGTAAAGTATTTGACAGATGCTCAAGAGATGTTCTGCAATGTTGGAGAAATTTGCTT 2040  
QY 2041 TAGTACTATCACTTCTGCTGCTCACTGATGATGACAAAGAGAAAGCTGTTGCATCACTGA 2100  
Db 2041 TAGTACTATCACTTCTGCTGCTCACTGATGATGACAAAGAGAAAGCTGTTGCATCACTGA 2100  
QY 2101 TACAGAGAAAGGAGCAAGCACTTTGAAACACAGCTGGGTTTCTTCTCAAGAGCTCCCA 2160  
Db 2101 TACAGAGAAAGGAGCAAGCACTTTGAAACACAGCTGGGTTTCTTCTCAAGAGCTCCCA 2160  
QY 2161 TCCAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2220  
Db 2161 TCCAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2220  
QY 2221 TGCA 2224  
Db 2221 TGCA 2224

RESULT 2  
AF262962 2411 bp mRNA linear PRI 30-MAY-2000  
LOCUS AF262962  
DEFINITION Homo sapiens VASA protein mRNA, complete cds.  
ACCESSION AF262962  
VERSION AF262962.1 GI:8102020  
KEYWORDS  
SOURCE  
ORGANISM  
Homo sapiens (human)  
Homo sapiens  
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE  
1 (bases 1 to 2411)  
Rocha, D. and Affara, N.  
TITLE Cloning and characterization of the human VASA gene  
JOURNAL Unpublished  
AUTHORS Rocha, D. and Affara, N.  
TITLE Direct Submission  
JOURNAL Submitted (02-MAY-2000) Department of Pathology, University of  
Cambridge, Tennis Court Road, Cambridge CB2 1QP, UK  
FEATURES  
source 1..2411

Query Match 90.0%; Score 2001; DB 9; Length 2411;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 2201; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
ORIGIN  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/chromosome="5"  
18..2192  
/codon\_start=1  
/product="VASA protein"  
/protein\_id="AA272705.1"  
/db\_xref="GI:8102021"  
/translation="MGDEMEAFILPHMSVYPIPEKORVSEGENDNFRRTASSEM  
DQPSRDRDFMKSGRPASGNFENRDRDGEKNDNTSTNGFVGKSPKRGFSNKE  
DGSSGFMRPSSNDCEDNPTNRNRSFKRGYVDGNNSESGYRGRGSGFRGCGF  
GLGSPNNDLDPCEMORTGLRPSRPVLSGNGDTSQSRSGSSESGYKGLNEV  
ITGSGKNSWKSSEAEAGSSTGQPRVTVIPPPEDDESIFAHYOTGIFDKYDTLV  
EYSGHDAFPAILTFEEANLCOTLNNNIAKAGYTKLTPQKYSIPILAGRDMAQOT  
GSGKTAFLPLILAHMHMDGITASRFKEILOBECTIIVPTRELVNOIYEAERKSGT  
CYRAVIVGCTGLGSIQIYVGCNIIICATPBRILDIIEKEIKQIKYIVDEDR  
MIDMGEPKMLISCPKPSRQOTLNFSTFPEBIOFLAEFLKSYLFAVQV  
GGACHDQVQLVQVQPSRREKVEILNRIIGERTMVEVETKKADFLATFLCQKIS  
TTSIHEDRQRREQLGDFRQKCPVIVASVARGDIEVQCHVIFNDLPSTIDEY  
VRIIGRTGRCNGTGRAISFDLESNDHIAQPLVKTLTAAQCDVPAMLEIARSTYIPG  
FSGSTRGNVFAVDTRKKSITLNTAFSSSQAPNVDESMD"

QY 20 GATGAAGATTGGAGAGCAGAAATCAACCTCATATGCTTCTATGTTCCATATTTAG 79  
Db 24 GATGAAGATTGGAGAGCAGAAATCAACCTCATATGCTTCTATGTTCCATATTTAG 83  
QY 80 AAGATAGGTATTTCTGGAGAAAATGAGACAAATTTTAAACAGAGCTCCAGCTTCATCATCA 139  
Db 84 AAGATAGGTATTTCTGGAGAAAATGAGACAAATTTTAAACAGAGCTCCAGCTTCATCATCA 143  
QY 140 GAAATGATGATGAGACCTTCTGAGAGATGATCAATTTGAAAGAGATTTGCTCTGGG 199  
Db 144 GAAATGATGATGAGACCTTCTGAGAGATGATCAATTTGAAAGAGATTTGCTCTGGG 203  
QY 200 CGAAATTTTGGAGAGAGATGCTGTGATGATTAATTAACGAGATTAATCATCCAAATG 259  
Db 204 CGAAATTTTGGAGAGAGATGCTGTGATGATTAATTAACGAGATTAATCATCCAAATG 263  
QY 260 GGTGTTTGGAGTTGAGAAAGGTTTGGAAACAGAGTTTTCAAACAGAGTTTGA 319  
Db 264 GGTGTTTGGAGTTGAGAAAGGTTTGGAAACAGAGTTTTCAAACAGAGTTTGA 323  
QY 320 GATGATGATGATGCTGTGTTTCTGAGAGAGTCTAGTAATGACTGGAGATTAATCAACA 379  
Db 324 GATGATGATGATGCTGTGTTTCTGAGAGAGTCTAGTAATGACTGGAGATTAATCAACA 383  
QY 380 CGAAACAGAGGTTTTCAGAGAGCGGCTATATGAGATGAGAAATTAATTCAGAACTTCA 439  
Db 384 CGAAACAGAGGTTTTCAGAGAGCGGCTATATGAGATGAGAAATTAATTCAGAACTTCA 443  
QY 440 GGGCCATACAGAGAGGTTGAGAGAGGTTGTTCCAGAGTCCCGTGGAGGATTTGCTGA 499  
Db 444 GGGCCATACAGAGAGGTTGAGAGAGGTTGTTCCAGAGTCCCGTGGAGGATTTGCTGA 503  
QY 500 GGAAGTCCAAATATGACTTATGACCCAGACAGATGATGACAGCGCATGCTGGGCTTTT 559  
Db 504 GGAAGTCCAAATATGACTTATGACCCAGACAGATGATGACAGCGCATGCTGGGCTTTT 563  
QY 560 GGTTCATAGAACCGCTTTTAAGTGGCCACAGTAATGCTATCTTCCAAAGCAGAAAGT 619  
Db 564 GGTTCATAGAACCGCTTTTAAGTGGCCACAGTAATGCTATCTTCCAAAGCAGAAAGT 623  
QY 620 GGCAGTGAAGTGAACGAGTGTGTTACAAAGGTTTAAATGAAGATTAATCAAGGCTCT 679  
Db 624 GGCAGTGAAGTGAACGAGTGTGTTACAAAGGTTTAAATGAAGATTAATCAAGGCTCT 683  
QY 680 GGAAGAAATTTCTTGAAGTCAAGAACGAGAGAGAGAAAGTATGATGATCTCAAGAGCA 739



Db 684 GGAAAGAAATCTTGGAAAGTCAGAAAGCAAGAGGAGAAAGTAGGATACCTAAGACCA 743  
 Qy 740 AAGTAGACCTAGTACCCCTCTCCACCTGAGGATGAGATCTCCATCTTTGACATTTAT 799  
 Db 744 AAGTAGACCTAGTACCCCTCTCCACCTGAGGATGAGATCTCCATCTTTGACATTTAT 803  
 Qy 800 CAGACAGGCAATAAATCTTGCACAAATACGACACTATTTCTTGGAAAGTGTCTGACATGAT 859  
 Db 804 CAGACAGGCAATAAATCTTGCACAAATACGACACTATTTCTTGGAAAGTGTCTGACATGAT 863  
 Qy 860 GACACACAGGCAATCTTGCACAAATCTTGCACAAATCTTGCACAAATCTTGCACAAAT 919  
 Db 864 GACACACAGGCAATCTTGCACAAATCTTGCACAAATCTTGCACAAATCTTGCACAAAT 923  
 Qy 920 ATTGCTAAAGCTGTTATATCTAAGCTTACTCTGTGCAAAAATACAGATTCCTATCAT 979  
 Db 924 ATTGCTAAAGCTGTTATATCTAAGCTTACTCTGTGCAAAAATACAGATTCCTATCAT 983  
 Qy 980 CTTCGACAGAGGATTTGATGCTTGTGCTCAAAACAGGCTCTGGGAAAGCTGGCTTTT 1039  
 Db 984 CTTCGACAGAGGATTTGATGCTTGTGCTCAAAACAGGCTCTGGGAAAGCTGGCTTTT 1043  
 Qy 1040 CTTCCTACCAATTTTGGCTCATATGATGATGATGATGATGATGATGATGATGATGAT 1099  
 Db 1044 CTTCCTACCAATTTTGGCTCATATGATGATGATGATGATGATGATGATGATGATGAT 1103  
 Qy 1100 TTGACAGAACCAAGATGATTTATTTGACACCACTGACAGATTTGATGATGATGAT 1159  
 Db 1104 TTGACAGAACCAAGATGATTTATTTGACACCACTGACAGATTTGATGATGATGAT 1163  
 Qy 1160 TTGAGAGCCAGAAATTTTCTTTGGGACTGTGTAGAGCTGTGTATATATATGAGGGA 1219  
 Db 1164 TTGAGAGCCAGAAATTTTCTTTGGGACTGTGTAGAGCTGTGTATATATATGAGGGA 1223  
 Qy 1220 ACCCAGCTGGACATTCATTCGACAAATGACAAAGCTGTATATATATATGAGGGA 1279  
 Db 1224 ACCCAGCTGGACATTCATTCGACAAATGACAAAGCTGTATATATATATGAGGGA 1283  
 Qy 1280 CTTCGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1339  
 Db 1284 CTTCGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1343  
 Qy 1340 TTGATTTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1399  
 Db 1344 TTGATTTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1403  
 Qy 1400 TTATTTTCTTGGCCAGGAATGCAATCAAGGAACAGGCGCAACCTTTATGTCAGTGA 1459  
 Db 1404 TTATTTTCTTGGCCAGGAATGCAATCAAGGAACAGGCGCAACCTTTATGTCAGTGA 1463  
 Qy 1460 ACTTTTCCAGAGAAATTTCAAGGTTGCTGACAGATTTTAAAGTCAATATCTGTTT 1519  
 Db 1464 ACTTTTCCAGAGAAATTTCAAGGTTGCTGACAGATTTTAAAGTCAATATCTGTTT 1523  
 Qy 1520 GTTGCTGTGACAGAGTGGTGGACATGTAGAGATTTTCAAGACCGTTCTCCAGATT 1579  
 Db 1524 GTTGCTGTGACAGAGTGGTGGACATGTAGAGATTTTCAAGACCGTTCTCCAGATT 1583  
 Qy 1580 GGGCAGTCTTCAAAAAGAAAGCTGTGGAATTTGGAACATATAGGGATGAAAGA 1639  
 Db 1584 GGGCAGTCTTCAAAAAGAAAGCTGTGGAATTTGGAACATATAGGGATGAAAGA 1643  
 Qy 1640 ACTATGCTCTTTGTTGAACCTAAGAAAAAGCAGATTTTATCTGCAACTTTCTTTGTC 1699  
 Db 1644 ACTATGCTCTTTGTTGAACCTAAGAAAAAGCAGATTTTATCTGCAACTTTCTTTGTC 1703  
 Qy 1700 GAAAAATATCAATCAAGATTCATGATGATGATGATGATGATGATGATGATGATGAT 1759  
 Db 1704 GAAAAATATCAATCAAGATTCATGATGATGATGATGATGATGATGATGATGATGAT 1763  
 Qy 1760 CTTCGAGATTTTCTTGGAAAGTGCAGATTTCTTGTTCCTTCACTGATGATGTCAGA 1819

Db 1764 CTTCGAGATTTTCTTGGAAAGTGCAGATTTCTTGTTCATCTTCAAGTGCAGAGA 1823  
 Qy 1820 GGGCTGATATTTGAAAAATGTGCAACATGTTATCAATTTTATCTTCTTACCATTTGAT 1879  
 Db 1824 GGGCTGATATTTGAAAAATGTGCAACATGTTATCAATTTTATCTTCTTACCATTTGAT 1883  
 Qy 1880 GAATATGATGATGATTTGGGCGGATGATGATGATGATGATGATGATGATGATGATGAT 1939  
 Db 1884 GAATATGATGATGATTTGGGCGGATGATGATGATGATGATGATGATGATGATGATGAT 1943  
 Qy 1940 TTTTTCATCTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1999  
 Db 1944 TTTTTCATCTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2003  
 Qy 2000 GCTTAAAGAGATGTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2059  
 Db 2004 GCTTAAAGAGATGTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2063  
 Qy 2060 TTCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2119  
 Db 2064 TTCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2123  
 Qy 2120 ACTTTGAAACACAGCTGGGTTTCTTCTTCAACAGCTCCCAATCCAGATGATGATGAT 2179  
 Db 2124 ACTTTGAAACACAGCTGGGTTTCTTCTTCAACAGCTCCCAATCCAGATGATGATGAT 2183  
 Qy 2180 TGGATTTAAAGCCAAACATCTTCAAGCTGTGTTGATGCA 2224  
 Db 2184 TGGATTTAAAGCCAAACATCTTCAAGCTGTGTTGATGCA 2228

RESULT 3  
 HSM802178  
 LOCUS  
 DEFINITION Homo sapiens mRNA; cdna DKFZp434B1122 (from clone DKFZp434B1122); partial cds.  
 ACCESSION AL137462  
 VERSION AL137462.1 GI:6808043  
 KEYWORDS  
 SOURCE  
 ORGANISM Homo sapiens (human)

REFERENCE  
 AUTHORS Blum, H., Bauesachs, S., Mewes, H. W., Gassenhuber, J. and Wiemann, S.  
 DIRECT SUBMISSION  
 JOURNAL Submitted (15-JAN-2000) MIPS, Am Klopferspitz 18a, D-82152 Martinsried, GERMANY

COMMENT  
 This clone (DKFZp434B1122) is available at the RZPD in Berlin.  
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at <http://www.mips.biochem.mpg.de/proj/cdna/>.  
 Location/Qualifiers

FEATURES  
 source  
 1..2189  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="DKFZp434B1122"  
 /tissue\_type="testis"  
 /clone\_id="434 (synonym: htes3). Vector report1; host DH10B; sites NotI + SalI"  
 /dev\_stage="adult"  
 1..2189  
 /gene="DKFZp434B1122"  
 <1..1910  
 /gene="DKFZp434B1122"  
 /note="strong similarity to testis specific Helicase (VISA

```
homolog1"
/codon_start=3
/product="hypothetical protein"
/protein_id="CAB70750.1"
/db_xref="GI:808044"
/db_xref="GOA:Q9NOI0"
/db_xref="Swiss-Prot:Q9NOI0"
/translation="SFGRNRSNRFEDSGSGFRSSNDCEDEPNFRNRSKRGY
RDNNSESGEYRNRGRGSRFCRGSGFGLSPNNDDLPDECMQGTGLFSRRLPVLSG
TGNDDTSQSRSGSSEKRGYKGLNEVITSGKMSKEAGSSDTPQPKVYLP
PPEDDESIPAHYQGTINFDPYDILVLEGHDPAPILTFEEMNLQTNANILAG
YTKLPVQKSIPIILAGRDIMACQTSQGTAFILPLIAPHNMIDITSRFEOE
PECTIVATRELAVNOIYLEARKSFSGTCAVAVIYGGTOLGHSIROIVOGCNILCAP
GRIMDIIGKEKIG:KOIKYLVDABMDLMGRGPEMKLISGCMPSKEKROTLMS
ATPPEIORLAEPILKSNYLFVANGQVGGACRDYQYVLYVQPSREKULVEILRNIG
DERTWVETKKADEFIATPLCEKISTISIHDEQREBQALGDFGKCEVLVAT
SVAARGLDIENVQHVINFDELSTIDEVHAIIGTRGCGNGRAISFEDLESDHMLQP
LVKYLDAQODVPAMLEIAFSTYIPGFSSTGNVPAVSVDTRKSKSTLTAGSSSO
APNVDDESMD"
2128..2133
/polya_signal
/gene="DKFZp434B1122"
2147
polya_site
/gene="DKFZp434B1122"
```

## ORIGIN

Query Match 78.3%; Score 1742; DB 9; Length 2189;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 1942; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
QY 279 AGAGTTTGAACAAGAGTTTTCACACAGAGTTTGAAGATGATGATGCTGGTT 338
Db 1 AGAGTTTGAACAAGAGTTTTCACACAGAGTTTGAAGATGATGATGCTGGTT 60
QY 339 TCTGAGAGAGTCAATGATGATGATGATGATGATGATGATGATGATGATGAT 338
Db 61 TCTGAGAGAGTCAATGATGATGATGATGATGATGATGATGATGATGATGAT 120
QY 399 AGAAGCGCGCTATGAGATGAGATTAATTCAGAGCTTCAGGCGCATACAGAGAGTG 458
Db 121 AGAAGCGCGCTATGAGATGAGATTAATTCAGAGCTTCAGGCGCATACAGAGAGTG 180
QY 459 GAAAGGTAGTTTCCGAGGTTGCCGTGAGAGATTTGGTCTGAGAGATCCAAATATGACT 518
Db 181 GAAAGGTAGTTTCCGAGGTTGCCGTGAGAGATTTGGTCTGAGAGATCCAAATATGACT 240
QY 519 TAGACCGAGAGATGATGAGCGCATGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 578
Db 241 TAGACCGAGAGATGATGAGCGCATGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 300
QY 579 TAAATGCGACAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 638
Db 301 TAAATGCGACAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 360
QY 639 GTGTTCAAAAGTTTAAATGAGAGATTAATCAGGCTCTGAGAAATATCTTGAAGT 698
Db 361 GTGTTCAAAAGTTTAAATGAGAGATTAATCAGGCTCTGAGAAATATCTTGAAGT 420
QY 699 CAGAAGCAGAAAGAGAGAAAGTATGATGATGATGATGATGATGATGATGATGATGAT 758
Db 421 CAGAAGCAGAAAGAGAGAAAGTATGATGATGATGATGATGATGATGATGATGATGAT 480
QY 759 CTCTCCACCTGAGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGAT 818
Db 481 CTCTCCACCTGAGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGAT 540
QY 819 ACAATATGACACTATTTCTGAGAGATGATGATGATGATGATGATGATGATGATGATGAT 878
Db 541 ACAATATGACACTATTTCTGAGAGATGATGATGATGATGATGATGATGATGATGATGAT 600
QY 879 CTTTGAAGAGAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 938
Db 601 CTTTGAAGAGAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 660
```

```
QY 939 CTAAAGTTACTCTGTGCAAAAATACAGTATTCATCATCTTGACAGAGCATTTGA 998
Db 661 CTAAAGTTACTCTGTGCAAAAATACAGTATTCATCATCTTGACAGAGCATTTGA 720
QY 999 TGGCTGTGCTCAAAACAGAGGTCGAGGAAGTACGCGGCTTTCTCCACCAATTTGGCT 1058
Db 721 TGGCTGTGCTCAAAACAGAGGTCGAGGAAGTACGCGGCTTTCTCCACCAATTTGGCT 780
QY 1059 ATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1118
Db 781 ATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
QY 1119 TTATGTAGCAACAACCTGAGAAATTTGCAACAATTTATTTGGAAGCCAGAAAATTTT 1178
Db 841 TTATGTAGCAACAACCTGAGAAATTTGCAACAATTTATTTGGAAGCCAGAAAATTTT 900
QY 1179 CTTTGGACCTTGTTAAGAGCTGTTGTTATATATGAGGGAACCCAGCTGGACATTCNA 1238
Db 901 CTTTGGACCTTGTTAAGAGCTGTTGTTATATATGAGGGAACCCAGCTGGACATTCNA 960
QY 1239 TTGCAAAATAGTACAGAGCTGATATATATATATATATATATATATATATATATAT 1298
Db 961 TTGCAAAATAGTACAGAGCTGATATATATATATATATATATATATATATATATAT 1020
QY 1299 TCATAGGCAAAAGAAATTTGCTCAAAACAGATCAAAATCTTATGTTGATGATGAAGTG 1358
Db 1021 TCATAGGCAAAAGAAATTTGCTCAAAACAGATCAAAATCTTATGTTGATGATGAAGTG 1080
QY 1359 ATGCAATGTTGATATGAGTTTGGTTTGGTCAAGAAATGAAGATTAATTTCTGCGAGAA 1418
Db 1081 ATGCAATGTTGATATGAGTTTGGTTTGGTCAAGAAATGAAGATTAATTTCTGCGAGAA 1140
QY 1419 TGGCAATCAAGGAACAGCGCCAAACCTTATGATGAGCACTTTCCAGAGAAATTC 1478
Db 1141 TGGCAATCAAGGAACAGCGCCAAACCTTATGATGAGCACTTTCCAGAGAAATTC 1200
QY 1479 AAAGTTGCTCAGAGTTTAAAGTCAAAATATCTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1538
Db 1201 AAAGTTGCTCAGAGTTTAAAGTCAAAATATCTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1260
QY 1539 GTGAGCATGATGAGATGTTCAAGCAACGTTCTCCAACTTGGCCAGTTCTCAAAAAGAG 1598
Db 1261 GTGAGCATGATGAGATGTTCAAGCAACGTTCTCCAACTTGGCCAGTTCTCAAAAAGAG 1320
QY 1599 AAAAGCTGTTGAATTTCTGCAAAACATAGGGGATGAAGAACTATGCTTTGTTGAAA 1658
Db 1321 AAAAGCTGTTGAATTTCTGCAAAACATAGGGGATGAAGAACTATGCTTTGTTGAAA 1380
QY 1659 CTAAAGAAAACAGATTTTACTGCACTTTTCTTTGCAAGAAAATATCACTACAA 1718
Db 1381 CTAAAGAAAACAGATTTTATGCACTTTTCTTTGCAAGAAAATATCACTACAA 1440
QY 1719 GTATCATGATGATGAGGAACAGAGAGCGGGAACAAGCTCTTGGAATTTGCGCTTG 1778
Db 1441 GTATCATGATGATGAGGAACAGAGAGCGGGAACAAGCTCTTGGAATTTGCGCTTG 1500
QY 1779 GAAAGTGCCAGTTCTGTTGCTACTTCAATGATGATGATGATGATGATGATGATGATGAT 1838
Db 1501 GAAAGTGCCAGTTCTGTTGCTACTTCAATGATGATGATGATGATGATGATGATGATGAT 1560
QY 1839 TGGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1898
Db 1561 TGGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1620
QY 1899 GGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1958
Db 1621 GGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680
QY 1959 ATAAATTTAGACAGCTCTAGTAAAGTATGACAGATGATCAACAGATGTTCTTG 2018
Db 1681 ATAAATTTAGACAGCTCTAGTAAAGTATGACAGATGATCAACAGATGTTCTTG 1740
QY 2019 CATGTTGAGAAATTTGCTTATGATACATATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2078
```

A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/JMML at: <http://image.lnl.gov>

Series: IRAX Plate: 91 Row: A Column: 10

This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 9507236.

Location/Qualifiers

```

1. .2363
   /organism="Homo sapiens"
   /mol_type="mRNA"
   /db_xref="taxon:9606"
   /clone="MGC:48304 IMAGE:529807"
   /issue_type="Testis"
   /clone_id="NH MGC_97"
   /lab_host="DH10B"
   /note="vector: pBluescript"
1. .2363
   /gene="DDX4"
   /note="synonym: VASA"
   /db_xref="LocusID:54514"
   /db_xref="MIM:605281"
38. .2110
   /codon_start=1
   /product="DDX4 protein"
   /protein_id="AAH7455.1"
   /db_xref="GI:28958131"
   /db_xref="LocusID:54514"
   /translation="MGDEWDWEAIEINPMSIYDIFEXDRYSGENGDNFRTPASSSEM
DDEPSRDHFMKSGFGRNFGADGECRKNRNTSMGFGYGSXGNFSNRSRE
DGSSGWRSSNDCEPNPTNRNGFSRGGNDIDPECMORTGGLFSRRPVLTGDN
GDTROSBSGSGSRGGYKGLNEBYITSGSKMSKSEBEGSESSDTCQYTIYIPPPP
EDBDSIPAHYQTSINFDKDTLIVESGHAAPALIFEEANLCQIINNVIAGAYTK
LTFVQKYSIPILILAGRLMAAQGTSGSKTAFLPLIAHMHGDTLSRKEIQEPCC
ILVAFRELVNKIYLEAKRSPFGTCVAAYIVGGTQVSHIRIQVGCNLICATPGL
MDIIGKEIKGIQKIRIYLDADRLMDMGPEMKLISCPGMPSEKQROTLMESATF
PEELIRAAEFELKSNYLFVAIVGGGACRDVQCTVLQVGFSGREKLELIRINDGR
TMTFVERKKADITATFLCOEKISTSHIGDRQREEOGLDPERKPCIVATSVAA
ARGKIDNVGVINFDFLSTIDEVHRIIGTCGNGTRALISFDLSDNHLAPLVKX
VLTDQAOVPAMLEELIAFSTYIIPFGSGTGNVAFSVTRKXSTLTNLAGSSQAPN
PVDESDN"
716. .1915
   /note="SmbB; Region: Superfamily II DNA and RNA helicases
DNA replication, recombination, and repair /
Transcription / Translation, ribosomal structure and
biogenesis"]
   /db_xref="CCD:COG0513"

```

ORIGIN

```

Query Match      63.4%; Score 1409; DB 9; Length 2363;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1709; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

misc_feature	misc_feature	misc_feature
510 ATATGACTTAGACCCAGACGAATGTATGACGCGACTGTGGCTTTGGTTCTTAGAA	569	
Db 432 ATATGACTTAGACCCAGACGAATGTATGACGCGACTGTGGCTTTGGTTCTTAGAA	491	
570 GACCAAGTATTAGTGGGACAGTTATGCTGTATCTTCTCAAGAGAGAGTGGCAGTGA	629	
Db 492 GACCAAGTATTAGTGGGACAGTATGCTGTATCTTCTCAAGAGAGAGTGGCAGTGA	551	
630 GTGAACAGAGTGTTCACAAAGTTTAATGAAGAGTAAATACAGGCTCTGGAAAGATT	689	
Db 552 GTGAACAGAGTGTTCACAAAGTTTAATGAAGAGTAAATACAGGCTCTGGAAAGATT	611	
690 CTTCGAAGTACGAAGACGAAGAGAGAGAAAGTAGTATCTTCAAGAGACCAAAATGACCT	749	
Db 612 CTTCGAAGTACGAAGACGAAGAGAGAGAAAGTAGTATCTTCAAGAGACCAAAATGACCT	671	
750 ACATACCCCTCTCCACCTCGAGGATAGAGATCCATCTTGCACATTATCAGACAGCA	809	
Db 672 ACATACCCCTCTCCACCTCGAGGATAGAGATCCATCTTGCACATTATCAGACAGCA	731	

QY 810 TAACTTGACAAATACGACACTATCTTGTGAGAGTGTGACATGATGACACACAG 869  
 Db 732 TAACTTGACAAATACGACACTATCTTGTGAGAGTGTGACATGATGACACACAG 791  
 QY 870 CAATTCGACTTTTGAGAGAGCTATCTCTGTGACACATGATATACAACTGCTTAAG 929  
 Db 792 CAATTCGACTTTTGAGAGAGCTATCTCTGTGACACATGATATACAACTGCTTAAG 851  
 QY 930 CTGCTTACTAAGCTTACTCTGTGACAAATAACAGATTCCTATCATCTTGACAGAC 989  
 Db 852 CTGCTTACTAAGCTTACTCTGTGACAAATAACAGATTCCTATCATCTTGACAGAC 911  
 QY 990 GAGATTGATGCTGTGCTGACAAACAGAGTGTGAGAAAGCTGCGGCTTTTCTCTACAA 1049  
 Db 912 GAGATTGATGCTGTGCTGACAAACAGAGTGTGAGAAAGCTGCGGCTTTTCTCTACAA 971  
 QY 1050 TTTGGCTCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1109  
 Db 972 TTTGGCTCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1031  
 QY 1110 CAGAGTATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1169  
 Db 1032 CAGAGTATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1091  
 QY 1170 GAAATTTTCTTTTGGGACTTGTGTAAAGCTGTTTATATATGATGAGGGAACCCAGCTGG 1229  
 Db 1092 GAAATTTTCTTTTGGGACTTGTGTAAAGCTGTTTATATATGATGAGGGAACCCAGCTGG 1151  
 QY 1230 GACATTCATTCGACAAATAGTACAGAGCTGATATATATATGATGATGATGATGATGATGAT 1289  
 Db 1152 GACATTCATTCGACAAATAGTACAGAGCTGATATATATATGATGATGATGATGATGATGAT 1211  
 QY 1290 TGATGATATCATAGGCAAGAAAGAAAGATGCTCTCAACAGATCAATATCTAGTTTGG 1349  
 Db 1212 TGATGATATCATAGGCAAGAAAGAAAGATGCTCTCAACAGATCAATATCTAGTTTGG 1271  
 QY 1350 ATGAGCTGATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1409  
 Db 1272 ATGAGCTGATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1331  
 QY 1410 GCCCAGGATGCCATCAAG 1469  
 Db 1332 GCCCAGGATGCCATCAAG 1391  
 QY 1470 AGGAAATTCAGAGGTTGGCTGACAGATTTTAAAGTCAATATATCTGTTGCTGCTGG 1529  
 Db 1392 AGGAAATTCAGAGGTTGGCTGACAGATTTTAAAGTCAATATATCTGTTGCTGCTGG 1451  
 QY 1530 GACAGTGGGTGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1589  
 Db 1452 GACAGTGGGTGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1511  
 QY 1590 CAAAAAGAGAAAGCTGCTGTAATTTCTGCGAAACATAGGAGATGAAAGAACTATGCTCT 1649  
 Db 1512 CAAAAAGAGAAAGCTGCTGTAATTTCTGCGAAACATAGGAGATGAAAGAACTATGCTCT 1571  
 QY 1650 TTGTTGAACCTAAGAAAGAGAGATTTTCTGCAATTTTCTTTGTCAGAAAGAAATAT 1709  
 Db 1572 TTGTTGAACCTAAGAAAGAGAGATTTTCTGCAATTTTCTTTGTCAGAAAGAAATAT 1631  
 QY 1710 CAATACAGATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1769  
 Db 1632 CAATACAGATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1691  
 QY 1770 TTGCTTTGAGAAAGTCCGAGTCTTGTGTCTACTCTAGTACTGCCAGAGGCTGGAAT 1829  
 Db 1692 TTGCTTTGAGAAAGTCCGAGTCTTGTGTCTACTCTAGTACTGCCAGAGGCTGGAAT 1751  
 QY 1830 TTGAAATTTGCAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1889  
 Db 1752 TTGAAATTTGCAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1811  
 QY 1890 ATGCAATTTGGGCGTACTGCTGCTGTTGTGGAATCTGCGAGAGCAATTTCTTTTGTGATC 1949

Db 1812 ATGCAATTTGGGCGTACTGCTGCTGTTGTGGAATCTGCGAGAGCAATTTCTTTTGTGATC 1871  
 QY 1950 TTGAATCGGATTAACCAATTTAGCAAGAGCTCTAGTAAAGATATTGAACAGTGTCAACAG 2009  
 Db 1872 TTGAATCGGATTAACCAATTTAGCAAGAGCTCTAGTAAAGATATTGAACAGTGTCAACAG 1931  
 QY 2010 ATGTTCTGCAATGTTGGAAGAAATTTGCTTTAGTACATCTGCTGCTCACTGATG 2069  
 Db 1932 ATGTTCTGCAATGTTGGAAGAAATTTGCTTTAGTACATCTGCTGCTCACTGATG 1991  
 QY 2070 GTACAGAGAAAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2129  
 Db 1992 GTACAGAGAAAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2051  
 QY 2130 CAGCTGGTCTTCTTCTTCAAGAGCTCCCAATCCAGTATGATGATGATGATGATGATGATGAT 2189  
 Db 2052 CAGCTGGTCTTCTTCTTCAAGAGCTCCCAATCCAGTATGATGATGATGATGATGATGATGATGAT 2111  
 QY 2190 GCCAAACATCCTTCAAGTCTGCTGTTGATGCA 2224  
 Db 2112 GCCAAACATCCTTCAAGTCTGCTGTTGATGCA 2146

RESULT 5  
 AX748127  
 LOCUS AX748127 1984 bp mRNA linear PAT 20-JUN-2003  
 DEFINITION Sequence 1652 from Patent EP1308459.  
 ACCESSION AX748127  
 VERSION AX748127.1 GI:32132515  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
 AUTHORS  
 1 Isogai, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S.,  
 Yamamoto, D., I., Isono, Y., Hio, Y., Otsuka, K., Nagai, K., Irie, R.,  
 Tamechika, I., Seki, N., Yoshikawa, T., Otsuka, M., Nagahari, K. and  
 Masuko, Y.

TITLE Full-length cDNA sequences  
 JOURNAL Patent: EP 1308459-A 1652 07-MAY-2003;  
 Helix Research Institute (JP); Research Association for  
 Biotechnology (JP)

FEATURES  
 source  
 1..1984  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"

ORIGIN  
 Query Match 57.9%; Score 1288; DB 6; length 1984;  
 Best Local Similarity 99.7%; Pred. No. 0;  
 Matches 1488; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 733 AGGACCAAAAGTACCTACATACCCCTCCCTCCAGCTGAGATGAGACTCATCTTTGCTG 792  
 Db 292 AGGACCAAAAGTACCTACATACCCCTCCCTCCAGCTGAGATGAGACTCATCTTTGCTG 351  
 QY 793 ACATTAACAGACAGGATTAACCTTGACAAATATACAGACTATCTTGTGAGAGTCTGCG 852  
 Db 352 ACATTAACAGACAGGATTAACCTTGACAAATATACAGACTATCTTGTGAGAGTCTGCG 411  
 QY 853 ACATTAACAGACAGGATTAACCTTGACAAATATACAGACTATCTTGTGAGAGTCTGCG 912  
 Db 412 ACATTAACAGACAGGATTAACCTTGACAAATATACAGACTATCTTGTGAGAGTCTGCG 471  
 QY 913 TACAAACATTTGGTAAAGCTGTTTATACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 972  
 Db 472 TACAAACATTTGGTAAAGCTGTTTATACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 531  
 QY 973 TATCATTTGCGAGAGAGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1032

Db 532 TATCATCTTGCAGAGAGAGATTGATGGCTTGCCTCAACAGGGTCTGGAGAGCTG 591  
 Qy 1033 GGGTTTCTCTCAACCAATTTTGGCTCATATGATGATGAAATACCTGCAGAGCTTT 1092  
 Db 592 GGGTTTCTCTCAACCAATTTTGGCTCATATGATGATGAAATACCTGCAGAGCTTT 651  
 Qy 1093 TAAAGATTGCAGAAACAGAGTGTATTAATTGAGCAACCACTCGAATTTGGTCAACA 1152  
 Db 652 TAAAGATTGCAGAAACAGAGTGTATTAATTGAGCAACCACTCGAATTTGGTCAACA 711  
 Qy 1153 GATTATTTTGAAGCCAGAAAATTTTCTTTGGGACTTGTGTATGAGCTGTGTATATA 1212  
 Db 712 GATTATTTTGAAGCCAGAAAATTTTCTTTGGGACTTGTGTATGAGCTGTGTATATA 771  
 Qy 1213 TGGGGGAAACCACTGGGACATTCATTCATTCAGCAATAGTACAGAGCTGTATATATATG 1272  
 Db 772 TGGGGGAAACCACTGGGACATTCATTCATTCAGCAATAGTACAGAGCTGTATATATG 831  
 Qy 1273 TGGTACTCTGGAAGACTGATGATATCATAGCCAAAGAAAAGATTTGTTCTCAACAGAT 1332  
 Db 832 TGGTACTCTGGAAGACTGATGATATCATAGCCAAAGAAAAGATTTGTTCTCAACAGAT 891  
 Qy 1333 CAATATCTTGTGGTGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1392  
 Db 892 CAATATCTTGTGGTGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 951  
 Qy 1393 GAAAGATTAAATTTCTGCCCAGAGAAATGCCATCAAGAGAAACAGCCCAACCTTATGTT 1452  
 Db 952 GAAAGATTAAATTTCTGCCCAGAGAAATGCCATCAAGAGAAACAGCCCAACCTTATGTT 1011  
 Qy 1453 CAGTGCACATTTTCCAGAGAAATTCAGAGAGTGGCTGAGAGAGTGGTAAAGTCAATTA 1512  
 Db 1012 CAGTGCACATTTTCCAGAGAAATTCAGAGAGTGGCTGAGAGAGTGGTAAAGTCAATTA 1071  
 Qy 1513 TCTGTTGTTGCTGTGGAACAAGTGGTGGAGATGAGATGATGATGATGATGATGATGAT 1572  
 Db 1072 TCTGTTGTTGCTGTGGAACAAGTGGTGGAGATGAGATGATGATGATGATGATGATGAT 1131  
 Qy 1573 CCAAGTTGGCAGTTCTCAAAAAGAGAAAGCTGCTGGAATTCCTCGAAACATAGGGGA 1632  
 Db 1132 CCAAGTTGGCAGTTCTCAAAAAGAGAAAGCTGCTGGAATTCCTCGAAACATAGGGGA 1191  
 Qy 1633 TGAAGAACTATGATGTTGTTGTAAGTAAAGAGATTTTACTGCAACTTTTCT 1692  
 Db 1192 TGAAGAACTATGATGTTGTTGTAAGTAAAGAGATTTTACTGCAACTTTTCT 1251  
 Qy 1693 TTGTCAAGAAAATATCAATCAAGTATCATGATGATGATGATGATGATGATGATGAT 1752  
 Db 1252 TTGTCAAGAAAATATCAATCAAGTATCATGATGATGATGATGATGATGATGATGAT 1311  
 Qy 1753 GCAAGCTCTGAGATTTTGGCTTGGAAAGTCCAGATTCCTGTTGCTACTTCACTAGTAC 1812  
 Db 1312 GCAAGCTCTGAGATTTTGGCTTGGAAAGTCCAGATTCCTGTTGCTACTTCACTAGTAC 1371  
 Qy 1813 TGGCAGAGGCTGGATATGAAAATGTCACAATGTTATCAATTTTATCTTCTCTAC 1872  
 Db 1372 TGGCAGAGGCTGGATATGAAAATGTCACAATGTTATCAATTTTATCTTCTCTAC 1431  
 Qy 1873 CATGATGATATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1932  
 Db 1432 CATGATGATATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1491  
 Qy 1933 AATTCTTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1992  
 Db 1492 AATTCTTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1551  
 Qy 1993 GACAGATGCTCAACAGAGATTTCTGCTGATGTTGGAAGAAATGCTTTTATGATCAT 2052  
 Db 1552 GACAGATGCTCAACAGAGATTTCTGCTGATGTTGGAAGAAATGCTTTTATGATCAT 1611  
 Qy 2053 TCCGAGCTTCAAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2112  
 Db 1612 TCCGAGCTTCAAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1671

Qy 2113 CAAGAGCACTTTGAAACACAGCTGGGTTTCTTCTTCTTCAAGAGCTCCCATTCAGTATGTA 2172  
 Db 1672 CAAGAGCACTTTGAAACACAGCTGGGTTTCTTCTTCTTCAAGAGCTCCCATTCAGTATGTA 1731  
 Qy 2173 TGAATCATGGGATTTAAAGCCAAAACATCTTCAAGTGTGTGTTTGAATGCA 2224  
 Db 1732 TGAATCATGGGATTTAAAGCCAAAACATCTTCAAGTGTGTGTTTGAATGCA 1783  
 RESULT 6  
 AK093439  
 LOCUS  
 DEFINITION  
 Homo sapiens cDNA FLJ36120 fls, clone TEST12022738, highly similar  
 to DEAD BOX PROTEIN 4.  
 ACCESSION  
 AK093439  
 KEYWORDS  
 oligo capping; fls (full insert sequence).  
 SOURCE  
 Homo sapiens (human)  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 AUTHORS  
 1 Nishi,T., Ota,T., Nakagawa,S., Senoh,A., Mizuguchi,H., Inagaki,H.,  
 Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,  
 Yamamoto,J., Isono,Y., Kawai,Hio,Y., Saito,K., Nishikawa,T.,  
 Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M.,  
 Kikuchi,H., Kanda,K., Wagatsuma,M., Mukakawa,K., Kanehori,K.,  
 Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y.,  
 Sugano,S., Nagahari,K., Masuno,Y., Nagai,X. and Iisogai,T.  
 NEDO human cDNA sequencing project  
 TITLE  
 NEDO human cDNA sequencing project  
 JOURNAL  
 Unpublished  
 REFERENCE  
 2 (bases 1 to 1984)  
 Iisogai,T. and Yamamoto,J.  
 Direct Submission  
 TITLE  
 Submitted (04-JUL-2002) Takao Iisogai, FLJ Project (HRI Team); 2-6-7  
 Kazusa-Kamatari, Kisarazu, Chiba 252-0812, Japan  
 (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)  
 COMMENT  
 NEDO human cDNA sequencing project supported by Ministry of  
 Economy, Trade and Industry of Japan; cDNA full insert sequencing:  
 Research Association for Biotechnology (RAB); cDNA library  
 construction: Helix Research Institute (HRI) (supported by Japan  
 Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,  
 HRI, and Biotechnology Center, National Institute of Technology and  
 Evaluation; clone selection for full insert sequencing: HRI and  
 RAB; annotation: HRI and RAB.  
 FEATURES  
 source  
 1. 1984  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="TEST12022738"  
 /feature\_type="testis"  
 /clone\_libs="TEST12"  
 /note="Cloning vector: pMB18SFL3"  
 ORIGIN  
 Query Match 57.9%; Score 1288; DB 9; Length 1984;  
 Best local similarity 99.7%; Pred. No. 0;  
 Matches 1488; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 Qy 733 AGAACCAAAAGTGAACCTTACATACCCCTCTCTCCAGCTGAGATGAGAGCTTCATCTTTGC 792  
 Db 292 AGAACCAAAAGTGAACCTTACATACCCCTCTCTCCAGCTGAGATGAGAGCTTCATCTTTGC 351  
 Qy 793 ACATTATGACAGAGCACTTAACTTGAACAATAGCACTATCTTGTGGAAGTGTGCG 852  
 Db 352 ACATTATGACAGAGCACTTAACTTGAACAATAGCACTATCTTGTGGAAGTGTGCG 411  
 Qy 853 ACATGATGACACACAGCAATCTGACTTTTGAAGAGCTATCTGTGTACAGACTGAA 912  
 Db 412 ACATGATGACACACAGCAATCTGACTTTTGAAGAGCTATCTGTGTGTACAGACTGAA 471

Qy	913	TAAACAATGCTAAACGCTGTTATPCTAGCTTACCTGTCGCAAAAAATACAGTATCC	972
Db	472	TAAACAATGCTAAACGCTGTTATPCTAGCTTACCTGTCGCAAAAAATACAGTATCC	531
Qy	973	TATCATACTTGCAGAGCAGATTTGATGCGCTTGCTCAAAACGAGTCTGGAAACATGC	1032
Db	532	TATCATACTTGCAGAGCAGATTTGATGCGCTTGCTCAAAACGAGGCTGGAAACATGC	591
Qy	1033	GGCTTTCTCCACCAATTTTGGCTCATATGATGCAATGGAATPACTGCCAGTCGTT	1092
Db	592	GGCTTTCTCCACCAATTTTGGCTCATATGATGCAATGGAATPACTGCCAGTCGTT	651
Qy	1093	TAAAGATTGACAGAACCAAGTGTATATTTGAGCACCAGCTCGAATTTGGTCAACA	1152
Db	652	TAAAGATTGACAGAACCAAGTGTATATTTGAGCACCAGCTCGAATTTGGTCAACA	711
Qy	1153	GATTTATTTGGAAGCCGAAATTTTCTTTGGGACTTGTPAAGCTGTGTTATATA	1212
Db	712	GATTTATTTGGAAGCCGAAATTTTCTTTGGGACTTGTPAAGCTGTGTTATATA	771
Qy	1213	TGGGGGAACCCAGCTGGGACATTCATTCGACAAATAGTACAGGCTGTATATATATG	1272
Db	772	TGGGGGAACCCAGCTGGGACATTCATTCGACAAATAGTACAGGCTGTATATATATG	831
Qy	1273	TGCTACTCTGGAAGACTGATGATATCATAGGCAAAAGAAAGTTGCTCAACAGAT	1332
Db	832	TGCTACTCTGGAAGACTGATGATATCATAGGCAAAAGAAAGTTGCTCAACAGAT	891
Qy	1333	CAAAATATTGATTTGGATGAACGATGATGAGTGTGATGAGTTTGGTCCGAAT	1392
Db	892	CAAAATATTGATTTGGATGAACGATGATGAGTGTGATGAGTTTGGTCCGAAT	951
Qy	1393	GAAAGATTATTTCTTCTCCGCAAGATGCCATCAAAAGAACGCGCAACCCCTATGTT	1452
Db	952	GAAAGATTATTTCTTCTCCGCAAGATGCCATCAAAAGAACGCGCAACCCCTATGTT	1011
Qy	1453	CAGTGCACTTTTCCAGAGAAATTCAAAGTTGCGTGCAGAGTTTAAAGTCAATTA	1512
Db	1012	CAGTGCACTTTTCCAGAGAAATTCAAAGTTGCGTGCAGAGTTTAAAGTCAATTA	1071
Qy	1513	TCTGTTTGCTGCTGTGGAACAAGTGGTGAAGATATGAGATGTTACAGCAACGTTCT	1572
Db	1072	TCTGTTTGCTGCTGTGGAACAAGTGGTGAAGATATGAGATGTTACAGCAACGTTCT	1131
Qy	1573	CCAAAGTGGCCAGTCTCAAAAAAGAAAAGCTCGTTGAATTTCTGCGAAACATAGGGGA	1632
Db	1132	CCAAAGTGGCCAGTCTCAAAAAAGAAAAGCTCGTTGAATTTCTGCGAAACATAGGGGA	1191
Qy	1633	TGAAGAATCATATGCTCTTTGTGAACCTAAGAAAAAGAGATTTTATCGCAACTTCT	1692
Db	1192	TGAAGAATCATATGCTCTTTGTGAACCTAAGAAAAAGAGATTTTATTCGCACTTTCT	1251
Qy	1693	TTCGCAAGAAAAATATCAACTTCAAGATATCCATGCTGATCGGAACAGAGACCGGGA	1752
Db	1252	TTCGCAAGAAAAATATCAACTTCAAGATATCCATGCTGATCGGAACAGAGACCGGGA	1311
Qy	1753	GCAAGCTCTGGAGATTTTCGCTTTGGAAAGTCCCAAGTCTTGTGCTACTCTAGTAGC	1812
Db	1312	GCAAGCTCTGGAGATTTTCGCTTTGGAAAGTCCCAAGTCTTGTGCTACTCTAGTAGC	1371
Qy	1813	TGCGAGAGGGCTGATATGAAATGTGCACATGTTATCAATTTTGAATCTTCTTAC	1872
Db	1372	TGCGAGAGGGCTGATATGAAATGTGCACATGTTATCAATTTTGAATCTTCTTAC	1431
Qy	1873	CATTGATGAAATATGCTCATCGAATTTGGGCGTACTGCTGTGGGAATACCTGCAGAC	1932
Db	1432	CATTGATGAAATATGCTCATCGAATTTGGGCGTACTGCTGTGGGAATACCTGCAGAC	1491
Qy	1933	AATTTCTTTTGGATCTTGAATCGGATTAACCATTTAGCAGACGCTCTAGTAAAGATT	1992
Db	1492	AATTTCTTTTGGATCTTGAATCGGATTAACCATTTAGCAGACGCTCTAGTAAAGATT	1551
Qy	1993	GACAGATGCTCAACAGAGATTTCTGTCATGCTTGGAAAGAAATTTGCTTATGATACAT	2052

Db	1552	GACAGATGCTCAACAGAGATTTCTGTCATGCTTGGAAAGAAATTTGCTTATGATACAT	1611
Qy	2053	TCCTGGCTTCAGTGGTAGTACAAAGAGAAACGCTGTTTGATCATGTTGATACAGAAAGG	2112
Db	1612	TCCTGGCTTCAGTGGTAGTACAAAGAGAAACGCTGTTTGATCATGTTGATACAGAAAGG	1671
Qy	2113	CAGAGACTTTGAAACACAGCTGGGTTTCTTCTTCAAGAGCTCCCAATCCAGTAGATGA	2172
Db	1672	CAGAGACTTTGAAACACAGCTGGGTTTCTTCTTCAAGAGCTCCCAATCCAGTAGATGA	1731
Qy	2173	TGAGCATGGGATTAAGCCAAACATCCTTCAAGTCTGTGTTTGATGCA	2224
Db	1732	TGAGCATGGGATTAAGCCAAACATCCTTCAAGTCTGTGTTTGATGCA	1783

RESULT 7	AR415286	504 bp	DNA	linear	PAT 18-DEC-2003
LOCUS	AR415286				
DEFINITION	Sequence 2923 from patent US 6639063.				
ACCESSION	AR415286				
VERSION	AR415286.1	GI:40170396			
KEYWORDS	Unknown.				
SOURCE	Unknown.				
ORGANISM	Unclassified.				
REFERENCE	1 (bases 1 to 504)				
AUTHORS	Edwards, J.-B.D.M., Robert S. and Giordano, J.-Y.				
TITLE	EST's and encoded human proteins				
JOURNAL	Patent: US 6639063-A 2923 28-OCT-2003;				
FEATURES	Location/Qualifiers				
source	1..504				
	/organism="unknown"				
	/mol_type="genomic DNA"				

ORIGIN					
Query Match	20.1%;	Score 448;	DB 6;	Length 504;	
Best Local Similarity	100.0%;	Pred. No. 4.ee-237;			
Matches 448;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	

Qy	775	TGAGGACCTCATCTTTGACCAATTAACAGAGGACATTAACCTTGACAAATACGACATAT	834
Db	1	TGAGGACCTCATCTTTGACCAATTAACAGAGGACATTAACCTTGACAAATACGACATAT	60
Qy	835	TCTTGGAAGTGTCTGACATGATGACACACAGCAATTCGACTTTTGAAGAAGCTAA	894
Db	61	TCTTGGAAGTGTCTGACATGATGACACACAGCAATTCGACTTTTGAAGAAGCTAA	120
Qy	895	TCTCTGTGACACGATTAACAATGCTAAAGCTGTTATCTAAGCTTACTCTCTGT	954
Db	121	TCTCTGTGACACGATTAACAATGCTAAAGCTGTTATCTAAGCTTACTCTCTGT	180
Qy	955	GCAAAAATCACTATTTCTATCATCTTGCAGAGCGAGATTTGATGCTGTGCTCAAC	1014
Db	181	GCAAAAATCACTATTTCTATCATCTTGCAGAGCGAGATTTGATGCTGTGCTCAAC	240
Qy	1015	AGGGTCTGGGAAGACTGGGCTTTTCTCTTCAACAAATTTGGCTCATATGATGATGAG	1074
Db	241	AGGGTCTGGGAAGACTGGGCTTTTCTCTTCAACAAATTTGGCTCATATGATGATGAG	300
Qy	1075	AATAACTGCACTGCTTTTAAAGAGTTGCAGAGAACAGAGTATATTTGAGACCAAC	1134
Db	301	AATAACTGCACTGCTTTTAAAGAGTTGCAGAGAACAGAGTATATTTGAGACCAAC	360
Qy	1135	TGAGGATTTGTCAACACAGATTTATTTGGAAGCCGAAAAATTTCTTTGGACTGTGT	1194
Db	361	TGAGGATTTGTCAACACAGATTTATTTGGAAGCCGAAAAATTTCTTTGGACTGTGT	420
Qy	1195	AAGAGCTGTGTTATATATGAGGGGAAC	1222
Db	421	AAGAGCTGTGTTATATATGAGGGGAAC	448

Db	1552	GACGAGTCTCACAGGATGTTCTCGATGGTTGGAAATATGCTTTAGTACATCAT	1611		
QY	2053	TCCTGGCTTCAGTGGTAGTCAAAAGAGAAACGTGTTGCATCGAGTTGATACCAAGAAAGG	2112		
Db	1612	TCCTGGCTTCAGTGGTAGTCAAAAGAGAAACGTGTTGCATCGAGTTGATACCAAGAAAGG	1671		
QY	2113	CAAGAGACCTTGAACACAGCTGGGTTTCTTCTTCCAGAGCTCCCAATCCAGATGTA	2172		
Db	1672	CAGAGACACTTGAACACAGCTGGGTTTCTTCTTCCAGAGCTCCCAATCCAGATGTA	1731		
QY	2173	TGAGTCATGGGATTTAAAGCCAAACATCCTTCAAGTCTGTGTTTGATCA	2224		
Db	1732	TGAGTCATGGGATTTAAAGCCAAACATCCTTCAAGTCTGTGTTTGATCA	1783		
RESULT 7	AR415286	504 bp	DNA	linear	PAT 18-DEC-2003
LOCUS	AR415286				
DEFINITION	Sequence 2923 from patent US 6639063.				
ACCESSION	AR415286				
VERSION	AR415286.1	GI:40170396			
KEYWORDS	Unknown.				
ORGANISM	Unknown.				
REFERENCE	Unclassified.				
AUTHORS	Edwards,J.-B.D.M., Jobert,S. and Giordano,J.-Y.				
TITLE	BEST's and encoded human proteins				
JOURNAL	Patent: US 6639063-A 2923 28-Oct-2003;				
FEATURES	Location/Qualifiers				
SOURCE	1..504				
ORIGIN	/organism="unknown"				
	/mol_type="genomic DNA"				
Query Match	20.1%; Score 448; DB 6; Length 504;				
Best Local Similarity	100.0%; Pred. No. 4.5e-237;				
Matches 448; Conservative	0; Mismatches 0; Indels 0; Gaps 0;				
QY	775	TGAGGATCCATCTTTGACATTATGACAGGACATTAACCTTGACAAATACGACATAT	834		
Db	1	TGAGGATCCATCTTTGACATTATGACAGGACATTAACCTTGACAAATACGACATAT	60		
QY	835	TCTTGTGAAAGTGTCTGACATGATGACACCGACAGCAATTTCTGACTTTTGAAGACTAA	894		
Db	61	TCTTGTGAAAGTGTCTGACATGATGACACCGACAGCAATTTCTGACTTTTGAAGACTAA	120		
QY	895	TTCCTGCAACACCTGAATAACAACATGGTAAAGCTGGTAACTACTAAGTACTCTCTGT	954		
Db	121	TTCCTGCAACACCTGAATAACAACATGGTAAAGCTGGTAACTACTAAGTACTCTCTGT	180		
QY	955	GCAAAATATACAGATTCCTATCATACTCTGACGACGAGATTTGATGGCTGTGCTCAAC	1014		
Db	181	GCAAAATATACAGATTCCTATCATACTCTGACGACGAGATTTGATGGCTGTGCTCAAC	240		
QY	1015	AGGCTCTGGGAAGACCTGGGCTTTTCTCTCAACAATTTTGGCTCATATGATGATGATGG	1074		
Db	241	AGGCTCTGGGAAGACCTGGGCTTTTCTCTCAACAATTTTGGCTCATATGATGATGATGG	300		
QY	1075	AATTAACCTGCAGTCGTTTAAAGATTGACGAAACCGAGATGTATTATTGTAGACCAAC	1134		
Db	301	AATTAACCTGCAGTCGTTTAAAGATTGACGAAACCGAGATGTATTATTGTAGACCAAC	360		
QY	1135	TCAGGAATTGTCAACCAATTTATTGGAAGCCAGAAAATTTTCTTTGGACTGTGT	1194		
Db	361	TCAGGAATTGTCAACCAATTTATTGGAAGCCAGAAAATTTTCTTTGGACTGTGT	420		
QY	1195	AAGAGCTGTGTATTATATATGGGGAAAC	1222		
Db	421	AAGAGCTGTGTATTATATATGGGGAAAC	448		



RESULT 8  
BD110839  
LOCUS BD110839 504 bp DNA linear PAT 18-SEP-2002  
DEFINITION EST and encoded human protein.  
ACCESSION BD110839  
VERSION BD110839.1 GI:23205657  
KEYWORDS JP 2002010789-A/2916.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 504)  
Edwards, J.B.D.W., Robert, S., and Giordano, J.E.  
EST and encoded human protein  
Patent: JP 2002010789-A 2916 15-JAN-2002;  
JOURNAL GENSER CORP  
COMMENT OS Homo sapiens (human)  
PN JP 2002010789-A/2916  
PD 15-JAN-2002 JP 2002080989  
PF 07-AUG-2000 JP 2002080989  
PR 05-AUG-1999 US 60/147499  
PI JEAN BAPTISTE DUVAS MILNE EDWARDS, SEVELIN JOBERT, JEAN EVE PI  
GIORDANO  
PC C12N15/09, C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, PC  
C12N1/21  
PC C12N5/10, C12P21/02, C12P21/08, C12Q1/68, C12N15/00, C12N5/00, PC  
C12N15/00  
CC EST and encoded human protein  
FH Location/Qualifiers  
FT Key 224..472.  
CD CDS  
FEATURES  
source  
1..504  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Query Match 20.1%; Score 448; DB 6; Length 504;  
Best Local Similarity 100.0%; Pred. No. 4,6e-237;  
Matches 448; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 775 TGAGAGCTCATCTTTGACATATACAGAGGACATTAAGTTGACAAATACGACATAT 834  
DB 1 TGAGAGCTCATCTTTGACATATACAGAGGACATTAAGTTGACAAATACGACATAT 60  
QY 835 TCTTGTGAGAGTGTGAGATATGACACGACATTTCTGATTTGAGAAAGCTAA 894  
DB 61 TCTTGTGAGAGTGTGAGATATGACACGACATTTCTGATTTGAGAAAGCTAA 120  
QY 895 TCTTGTGAGAGTGTGAGATATGACACGACATTTCTGATTTGAGAAAGCTAA 954  
DB 121 TCTTGTGAGAGTGTGAGATATGACACGACATTTCTGATTTGAGAAAGCTAA 180  
QY 955 GCAAAATACAGATTTCTTCACTTGTGAGAGAGATTTGATGAGCTTGTGCTAAAC 1014  
DB 181 GCAAAATACAGATTTCTTCACTTGTGAGAGAGATTTGATGAGCTTGTGCTAAAC 240  
QY 1015 AGGCTGAGAGAGAGCTGCTTTCTTCTACCAATTTGGCTCATATGATGATGATG 1074  
DB 241 AGGCTGAGAGAGAGCTGCTTTCTTCTACCAATTTGGCTCATATGATGATGATG 300  
QY 1075 AATTAAGTGTGATTTTAAAGATTTGAGAGAGAGATTTGATGATGATGATG 1134  
DB 301 AATTAAGTGTGATTTTAAAGATTTGAGAGAGAGATTTGATGATGATGATG 360  
QY 1135 TCGAGATTTGATTTTAAAGATTTTAAAGATTTTAAAGATTTTAAAGATTTTAAAG 1194  
DB 361 TCGAGATTTGATTTTAAAGATTTTAAAGATTTTAAAGATTTTAAAGATTTTAAAG 420  
QY 1195 AAGAGCTGTGATTTTAAAGATTTTAAAGATTTTAAAGATTTTAAAGATTTTAAAG 1222  
DB 421 AAGAGCTGTGATTTTAAAGATTTTAAAGATTTTAAAGATTTTAAAGATTTTAAAG 448

RESULT 9  
AC008914  
LOCUS AC008914 139677 bp DNA linear PRI 19-APR-2002  
DEFINITION Homo sapiens chromosome 5 clone CTD-2270116, complete sequence.  
ACCESSION AC008914  
VERSION AC008914.7 GI:20198513  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 139677)  
DOE Joint Genome Institute and Stanford Human Genome Center.  
JOURNAL Direct Submission  
REFERENCE  
2 (bases 1 to 139677)  
DOE Joint Genome Institute.  
JOURNAL Direct Submission  
REFERENCE  
3 (bases 1 to 139677)  
DOE Joint Genome Institute and Stanford Human Genome Center.  
JOURNAL Direct Submission  
REFERENCE  
4 (bases 1 to 139677)  
DOE Joint Genome Institute and Stanford Human Genome Center.  
JOURNAL Direct Submission  
REFERENCE  
Submitted (19-APR-2002) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA  
On Apr 19, 2002 this sequence version replaced gi:19747123.  
Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center  
www.sngc.stanford.edu  
Quality: Phrap Quality >=40 99.8% of Sequence;  
Estimated Total Number of Errors is 0.2.  
FEATURES  
source  
1..139677  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="CTD-2270116"  
ORIGIN  
Query Match 12.3%; Score 273; DB 9; Length 139677;  
Best Local Similarity 100.0%; Pred. No. 9,4e-140;  
Matches 273; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1728 GTGATCGGGAACAGAGAGAGCGGAGCAAGCTTTGAGATTTTGGTTGAAAGTGC 1787  
DB 13765 GTGATCGGGAACAGAGAGAGCGGAGCAAGCTTTGAGATTTTGGTTGAAAGTGC 13824  
QY 1788 CAGTTCTTGTGCTACTTCACTAGCTGCGAGAGGCTGATATTTGAAAGTGC 1847  
DB 13825 CAGTTCTTGTGCTACTTCACTAGCTGCGAGAGGCTGATATTTGAAAGTGC 13884  
QY 1848 TTATCAATTTGATCTTCTTCACTAGCTGCGAGAGGCTGATATTTGAAAGTGC 1907  
DB 13885 TTATCAATTTGATCTTCTTCACTAGCTGCGAGAGGCTGATATTTGAAAGTGC 13944  
QY 1908 GTGCTGTGGAAATACGAGAGCAATTTCTTTTGAATCTTGAATCGGATTAACATT 1967  
DB 13945 GTGCTGTGGAAATACGAGAGCAATTTCTTTTGAATCTTGAATCGGATTAACATT 14004  
QY 1968 TAGCAGAGCTCTAGTAAAGTATGACAGATG 2000  
DB 14005 TAGCAGAGCTCTAGTAAAGTATGACAGATG 14037

RESULT 10



AC016632/c  
LOCUS AC016632 177185 bp DNA linear PRI 01-DEC-2002  
DEFINITION Homo sapiens chromosome 5 clone RP11-175M2, complete sequence.  
ACCESSION AC016632  
VERSION AC016632.7 GI:25989044  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
AUTHORS 1 (bases 1 to 177185)  
TITLE DOE Joint Genome Institute and Stanford Human Genome Center.  
JOURNAL Direct Submission  
REFERENCE 2 (bases 1 to 177185)  
AUTHORS Unpublished  
TITLE DOE Joint Genome Institute.  
JOURNAL Direct Submission  
REFERENCE 3 (bases 1 to 177185)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
JOURNAL Submitted (04-DEC-1999) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
REFERENCE 3 (bases 1 to 177185)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
JOURNAL Submitted (01-DEC-2002) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA  
COMMENT On Dec 1, 2002 this sequence version replaced gi:15290348.  
Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center  
www.sngc.stanford.edu  
Quality: Phrap Quality >=40 99.4% of Sequence;  
Estimated Total Number of Errors is 0.7.  
FEATURES  
source  
1. 177185  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="RP11-175M2"  
ORIGIN  
Query Match 12.3%; Score 273; DB 9; Length 177185;  
Best Local Similarity 100.0%; Pred. No. 9.3e-140;  
Matches 273; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1728 GTGATCGGGAACAGAGAGAGGAGAGAGCTTTGGAGATTTTCGTTGGAAAGGCC 1787  
DB 34889 GTGATCGGGAACAGAGAGAGGAGAGAGCTTTGGAGATTTTCGTTGGAAAGGCC 34830  
QY 1788 CAGTTCTTGTGCTACTCTCACTAGTCCAGAGAGGCTGATATTTGAAATGTCACATG 1847  
DB 34829 CAGTTCTTGTGCTACTCTCACTAGTCCAGAGAGGCTGATATTTGAAATGTCACATG 34770  
QY 1848 TTATCAATTTGATCTTCTTCTTACCATTTGATGATATTTGTCATCATGATGGCGGATCTG 1907  
DB 34769 TTATCAATTTGATCTTCTTCTTACCATTTGATGATATTTGTCATCATGATGGCGGATCTG 34710  
QY 1908 GTGCTTGTGGGAATAGTACTGAGAGAGCATTTTCCTTTTGTGATCTGATCGATACCATT 1967  
DB 34709 GTGCTTGTGGGAATAGTACTGAGAGAGCATTTTCCTTTTGTGATCTGATCGATACCATT 34650  
QY 1968 TAGCACAGCCTCTTAGTAAAGTATTTGACAGATG 2000  
DB 34649 TAGCACAGCCTCTTAGTAAAGTATTTGACAGATG 34617  
RESULT 11  
LOCUS AC016639 182126 bp DNA linear HTG 18-JUL-2000  
DEFINITION Homo sapiens chromosome 5 clone RP11-412L4, WORKING DRAFT SEQUENCE,  
7 ordered pieces.  
ACCESSION AC016639  
VERSION AC016639.5 GI:7711585  
KEYWORDS HTG; HTGS\_PHASE2; HTGS\_DRAFT.

SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
AUTHORS 1 (bases 1 to 182126)  
TITLE DOE Joint Genome Institute.  
JOURNAL Sequencing of Human Chromosome 5  
REFERENCE 2 (bases 1 to 182126)  
AUTHORS Unpublished  
TITLE DOE Joint Genome Institute.  
JOURNAL Direct Submission  
REFERENCE 2 (bases 1 to 182126)  
AUTHORS Submitted (04-DEC-1999) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
COMMENT On May 6, 2000 this sequence version replaced gi:7710162.  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov  
-----  
Project Information  
Center Project Name: 564841  
Center clone name: RP11-412L4  
-----  
Summary Statistics  
Consensus quality: 175214 bases at least Q40  
Consensus quality: 180404 bases at least Q30  
Consensus quality: 181112 bases at least Q20  
Estimated insert size: 178000; pulse-field gel estimation  
Estimated insert size: 181876; sum-of-contigs estimation  
Quality coverage: 6.41 in Q20 bases; pulse-field gel estimation  
Quality coverage: 6.26 in Q20 bases; sum-of-contigs estimation.  
NOTE: This is a 'working draft' sequence. It currently  
\* consists of 7 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 1 48575: contig of 48575 bp in length  
\* 48576 48575: gap of unknown length  
\* 48676 62810: contig of 1435 bp in length  
\* 62811 62910: gap of unknown length  
\* 62911 81913: contig of 19003 bp in length  
\* 81914 82013: gap of unknown length  
\* 82014 92166: contig of 10153 bp in length  
\* 92167 92266: gap of unknown length  
\* 92267 167478: contig of 75212 bp in length  
\* 167479 167579: gap of unknown length  
\* 167579 178445: contig of 10867 bp in length  
\* 178446 178545: gap of unknown length  
\* 178546 182126: contig of 3581 bp in length.  
FEATURES  
source  
1. 182126  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="RP11-412L4"  
/clone\_lib="RP11 human BAC library 11"  
ORIGIN  
Query Match 12.3%; Score 273; DB 2; Length 182126;  
Best Local Similarity 100.0%; Pred. No. 9.3e-140;  
Matches 273; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1728 GTGATCGGGAACAGAGAGAGGAGAGAGCTTTGGAGATTTTCGTTGGAAAGGCC 1787  
DB 36792 GTGATCGGGAACAGAGAGAGGAGAGAGCTTTGGAGATTTTCGTTGGAAAGGCC 36733  
QY 1788 CAGTTCTTGTGCTACTCTCACTAGTCCAGAGAGGCTGATATTTGAAATGTCACATG 1847

Db 36732 CAGTCTGTTGCTACTTACGATGCTGCCAGAGCGCTGATATTTGAAATTCGCAACATG 36673  
Qy 1848 TTATCATTTTGAATCTCTCTTACCATGATGATATGTCATCGAATTCGGCCGACTG 1907  
Db 36672 TTATCATTTTGAATCTCTCTTACCATGATGATATGTCATCGAATTCGGCCGACTG 36613  
Qy 1908 GTGCTGTGGGATTCGAGAGAGCAATTCCTTTTTCATCTGATCGGATTAACCAT 1967  
Db 36612 GTGCTGTGGGATTCGAGAGAGCAATTCCTTTTTCATCTGATCGGATTAACCAT 36553  
Qy 1968 TAGACAGCCTCTAGTAAAGATTTGACAGATG 2000  
Db 36552 TAGACAGCCTCTAGTAAAGATTTGACAGATG 36520  
RESULT 12  
AC022265 204250 bp DNA linear HTG 13-AUG-2000  
LOCUS AC022265  
DEFINITION Homo sapiens chromosome 5 clone RP11-332C17, WORKING DRAFT  
AC022265  
AC022265.3 GI:9719812  
VERSION AC022265.3  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 204250)  
AUTHORS Waterston,R.H.  
TITLE The sequence of Homo sapiens clone  
JOURNAL Unpublished  
2 (bases 1 to 204250)  
AUTHORS Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (27-JAN-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
COMMENT On Aug 7, 2000 this sequence version replaced gi:7940395.  
----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu/gsc/index.shtml  
----- Project Information -----  
Center project name: H NH0332C17  
----- Summary Statistics -----  
Sequencing vector: plasmid; 13k  
Chemistry: Dye-terminator Big Dye; 15k of reads  
Chemistry: Dye-terminator Big Dye; 15k of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 196499 bases at least Q40  
Consensus quality: 199226 bases at least Q30  
Consensus quality: 200804 bases at least Q20  
Insert size: 205000; agarose-fp  
Insert size: 202750; sum-of-contigs  
Quality coverage: 4.50 in Q20 bases; sum-of-contigs  
Quality coverage: 4.65 in Q20 bases; sum-of-contigs  
----- NOTE: This is a 'working draft' sequence. It currently  
\* consists of 16 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\*  
\* 1 1062: contig of 1062 bp in length  
\* 1063 1162: gap of unknown length  
\* 1163 2536: contig of 1374 bp in length  
\* 2537 2636: gap of unknown length  
\* 2637 5791: contig of 3155 bp in length  
\* 5792 5891: gap of unknown length

FEATURES  
Source \*  
\* 5892 10946: contig of 5055 bp in length  
\* 10947 11046: gap of unknown length  
\* 11047 15725: contig of 4679 bp in length  
\* 15726 15825: gap of unknown length  
\* 15826 21169: contig of 5344 bp in length  
\* 21170 21269: gap of unknown length  
\* 21270 27800: contig of 6531 bp in length  
\* 27801 27900: gap of unknown length  
\* 27901 33277: contig of 5376 bp in length  
\* 33277 33377: gap of unknown length  
\* 33377 42711: contig of 9335 bp in length  
\* 42711 42811: gap of unknown length  
\* 42811 56046: contig of 13235 bp in length  
\* 56047 56147: gap of unknown length  
\* 56147 77264: contig of 21117 bp in length  
\* 77264 77364: gap of unknown length  
\* 77364 97974: contig of 20611 bp in length  
\* 97974 98075: gap of unknown length  
\* 98075 122272: contig of 24198 bp in length  
\* 122272 122373: gap of unknown length  
\* 122373 148199: contig of 25827 bp in length  
\* 148200 148299: gap of unknown length  
\* 148300 175035: contig of 26736 bp in length  
\* 175036 175135: gap of unknown length  
\* 175136 204250: contig of 29115 bp in length.  
Location/Qualifiers  
1. .204250  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="RP11-332C17"  
1. .1062  
/note="assembly\_name:Contig12"  
1163. .2536  
/note="assembly\_name:Contig13"  
2637. .5791  
/note="assembly\_name:Contig14"  
5892. .10946  
/note="assembly\_name:Contig15"  
11047. .15725  
/note="assembly\_name:Contig16"  
15826. .21169  
/note="assembly\_name:Contig17"  
21270. .27800  
/note="assembly\_name:Contig18"  
27901. .33276  
/note="assembly\_name:Contig19"  
33377. .42711  
/note="assembly\_name:Contig20"  
clone\_end:17  
vector\_side:right"  
42812. .56046  
/note="assembly\_name:Contig21"  
56147. .77263  
/note="assembly\_name:Contig22"  
77364. .97974  
/note="assembly\_name:Contig23"  
98075. .122272  
/note="assembly\_name:Contig24"  
122373. .148199  
/note="assembly\_name:Contig25"  
148300. .175035  
/note="assembly\_name:Contig26"  
175136. .204250  
/note="assembly\_name:Contig27"  
ORIGIN  
Query Match 12.3%; Score 273; DB 2; Length 204250;  
Best Local Similarity 100.0%; Pred. No. 9.2e-140; Indels 0;  
Matches 273; Conservative 0; Mismatches 0; Gaps 0;  
Qy 1728 GTGATCGGGAACAAGAGAGCGGAGCAAGCTTTGGAGATTTTCCTTGGAAAGTCC 1787

Db 21580 GTGATCGGAACAGAGAGAGCGGAGAGAGCTCTTGAGAGATTTTCGTTTGAAAGGCC 21639

Qy 1788 CAGTCTCTTGTGCTACTTCACTTACCTGCGACAGGGCTGGATATTGAAATGTGCACATG 1847

Db 21640 CAGTCTCTTGTGCTACTTCACTTACCTGCGACAGGGCTGGATATTGAAATGTGCACATG 21699

Qy 1848 TTATCAATTTGATCTTCTCTCTACATGTAATGTTCATCGAATTGGGGCTACTG 1907

Db 21700 TTATCAATTTGATCTTCTCTCTACATGTAATGTTCATCGAATTGGGGCTACTG 21759

Qy 1908 GTCTGTGGGAATACCTGCGACAGCAATTTCTTTTGAATCTTGAATCGAATACCAT 1967

Db 21760 GTCTGTGGGAATACCTGCGACAGCAATTTCTTTTGAATCTTGAATCGAATACCAT 21819

Qy 1968 TAGCACAGCCTCTAGTAAAGTATTGACAGATG 2000

Db 21820 TAGCACAGCCTCTAGTAAAGTATTGACAGATG 21852

RESULT 13

AC022265/c

LOCUS

DEFINITION Homo sapiens chromosome 5 clone RP11-332C17, WORKING DRAFT

ACCESSION AC022265

VERSION AC022265.3 GI:9719812

KEYWORDS HTG; HTGS PHASE1; HTGS\_DRAFT.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 204250)

AUTHORS Waterston,R.H.

TITLE The sequence of Homo sapiens clone

JOURNAL Unpublished

2 (bases 1 to 204250)

Waterston,R.H.

Direct Submission

Submitted (27-JAN-2000) Genome Sequencing Center, Washington University School of Medicine, 444 Forest Park Parkway, St. Louis, MO 63108, USA

COMMENT On Aug 7, 2000 this sequence version replaced gi:7940395.

----- Genome Center -----

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc/index.shtml>

----- Project Information -----

Center project name: H NH0332C17

----- Summary Statistics -----

Sequencing vector: M13; 87%

Sequencing vector: plasmid; 13%

Chemistry: Dye-primer ET; 85% of reads

Chemistry: Dye-terminator Big Dye; 15% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 19649 bases at least Q40

Consensus quality: 200804 bases at least Q30

Consensus quality: 19926 bases at least Q20

Insert size: 205000; agarose-fp

Insert size: 202750; sum-of-ctnigs

Quality coverage: 4.50 in Q20 bases; agarose-fp

Quality coverage: 4.65 in Q20 bases; sum-of-ctnigs

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 16 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

1 1062: contig of 1062 bp in length

FEATURES

source

1..1063

1162: gap of unknown length

2536: contig of 1374 bp in length

2537

2636: gap of unknown length

2637

5791: contig of 315 bp in length

5792

5891: gap of unknown length

5892

10946: contig of 5055 bp in length

10947

11046: gap of unknown length

11047

15725: contig of 4679 bp in length

15726

15825: gap of unknown length

15826

21693: contig of 5344 bp in length

21694

21699: gap of unknown length

21700

27800: contig of 6531 bp in length

27801

33276: contig of 5376 bp in length

33277

33376: gap of unknown length

33377

42711: contig of 9335 bp in length

42712

42811: gap of unknown length

42812

56046: contig of 13235 bp in length

56047

56146: gap of unknown length

56147

77263: contig of 21117 bp in length

77264

77363: gap of unknown length

77364

97974: contig of 20611 bp in length

97975

98074: gap of unknown length

98075

122272: contig of 24198 bp in length

122273

122372: gap of unknown length

122373

148199: contig of 25827 bp in length

148200

148299: gap of unknown length

148300

175035: contig of 26736 bp in length

175036

175135: gap of unknown length

175136

204250: contig of 29115 bp in length.

Location/Qualifiers

1..204250

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="5"

/clone="RP11-332C17"

1..1062

/note="assembly\_name:Contig12"

1163..2536

/note="assembly\_name:Contig13"

2637..5791

/note="assembly\_name:Contig14"

5892..10946

/note="assembly\_name:Contig15"

11047..15725

/note="assembly\_name:Contig16"

15826..21169

/note="assembly\_name:Contig17"

21270..27800

/note="assembly\_name:Contig18"

27901..33276

/note="assembly\_name:Contig19"

33377..42711

/note="assembly\_name:Contig20

clone\_end:17

vector\_side:right"

42812..56046

/note="assembly\_name:Contig21"

56147..77263

/note="assembly\_name:Contig22"

77364..97974

/note="assembly\_name:Contig23"

98075..122272

/note="assembly\_name:Contig24"

122373..148199

/note="assembly\_name:Contig25"

148300..175035

/note="assembly\_name:Contig26"

175136..204250

/note="assembly\_name:Contig27"

ORIGIN

Query Match 7.8%; Score 174; DB 2; Length 204250;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-84;  
 Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1312 AAGATTGGTCTCAACAGATCAATCTAGTTTGGATGAAGTATGATGATGGA 1371  
 170192 AAGATTGGTCTCAACAGATCAATCTAGTTTGGATGAAGTATGATGATGGA 170133

QY 1372 TATGGTTTGGTCCGAATGAAGTAAATTTCTTGGCCAGAAATCCATCAAGA 1431  
 170132 TATGGTTTGGTCCGAATGAAGTAAATTTCTTGGCCAGAAATCCATCAAGA 170073

DB 1432 ACAGCCGCAACCTTATGTCGACACTTTCAGAGAAATCAAGGTT 1485  
 170072 ACAGCCGCAACCTTATGTCGACACTTTCAGAGAAATCAAGGTT 170019

QY 1432 ACAGCCGCAACCTTATGTCGACACTTTCAGAGAAATCAAGGTT 1485  
 170072 ACAGCCGCAACCTTATGTCGACACTTTCAGAGAAATCAAGGTT 170019

RESULT 14  
 AY100475 524 bp DNA linear NAM 15-JUN-2002  
 LOCUS Equus caballus VASA-like protein gene, partial sequence.  
 DEFINITION AY100475  
 VERSION AY100475.1 GI:21429223  
 KEYWORDS  
 SOURCE Equus caballus (horse)  
 ORGANISM Equus caballus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Perissodactyla; Equidae; Equus.  
 REFERENCE 1 (bases 1 to 524)  
 Woods, B.G., Ginther, O.J., Wentworth, A., Wentworth, B. and  
 Wiltbank, M.  
 TITLE Equine VASA Homolog  
 JOURNAL Unpublished  
 2 (bases 1 to 524)  
 Woods, B.G., Ginther, O.J., Wentworth, A., Wentworth, B. and  
 Wiltbank, M.  
 AUTHORS  
 DIRECT SUBMISSION  
 JOURNAL Submitted (30-APR-2002) Animal Health and Biomedical Sciences,  
 University of Wisconsin-Madison, 1656 Linden Drive, Madison, WI  
 53706, USA  
 FEATURES  
 source  
 1..524  
 location/Qualifiers  
 /organism="Equus caballus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9796"  
 <1..524  
 /note="similar to Drosophila melanogaster VASA protein"  
 ORIGIN  
 misc\_feature  
 Query Match 2.6%; Score 58; DB 4; Length 524;  
 Best Local Similarity 100.0%; Pred. No. 7.5e-20;  
 Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1997 GATGCTCAACAGATGTTCTGATGATGATGGAAGAAATGCTTTAGTACATCATTC 2054  
 455 GATGCTCAACAGATGTTCTGATGATGATGGAAGAAATGCTTTAGTACATCATTC 512

DB 211429 bp DNA linear HTG 22-FEB-2003  
 AC109144  
 LOCUS Mus musculus clone RP23-105P19, WORKING DRAFT SEQUENCE, 24  
 DEFINITION  
 unoriented pieces.  
 AC109144  
 VERSION AC109144.3 GI:28467368  
 HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 KEYWORDS Mus musculus (house mouse)  
 SOURCE Mus musculus  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 211429)  
 Birren, B., Nusbaum, C. and Lander, E.  
 AUTHORS Mus musculus, clone RP23-105P19

JOURNAL  
REFERENCES  
AUTHORS

Unpublished  
 2 (bases 1 to 211429)  
 Birren, B., Linton, J., Nusbaum, C., Lander, E., Allen, N.,  
 Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhalter, B.,  
 Brown, A., Camarata, J., Campiano, A., Chang, J., Chararo, B.,  
 Chepel, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A.,  
 Cooke, P., Dearellano, K., Dewar, J., Diaz, J.S., Dodge, S., Faro, S.,  
 Ferreira, P., Fitzhugh, N., Gage, D., Galagan, J., Gardyna, S.,  
 Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
 Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
 Kanat, A., Karatas, A., Kells, C., Lachoque, K., Lamas, R.,  
 Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C.,  
 Macdonald, P., Major, J., Margulis, N., Matthews, C., McCarthy, M.,  
 McKean, P., McKernan, K., Meldrum, J., Menus, L., Mihova, T.,  
 Mierga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,  
 Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,  
 Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,  
 Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, T.,  
 Roselli, M., Roy, A., Santos, R., Schuer, S., Schnupack, R., Seaman, S.,  
 Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
 Straus, N., Subramanian, A., Talmas, J., Testaye, S., Theodore, J.,  
 Totham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H.,  
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,  
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE  
JOURNAL  
REFERENCES  
AUTHORS

Direct Submission  
 Submitted (03-FEB-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 211429)  
 Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,  
 Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T.,  
 Boguslavsky, L., Boukhalter, B., Camarata, J., Chang, J., Chepel, Y.,  
 Collamore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K.,  
 Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,  
 Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S.,  
 Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,  
 Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
 Kanat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,  
 Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C.,  
 Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,  
 Meldrum, J., Menus, L., Mihova, T., Mierga, V., Murphy, T., Naylor, J.,  
 Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,  
 O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,  
 Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,  
 Roman, J., Schauer, S., Schnupack, R., Seaman, S., Severy, P., Smith, C.,  
 Spencer, B., Stange-Thomann, N., Stojanovic, N., Strubbs, M.,  
 Talamas, J., Testaye, S., Theodore, J., Totham, K., Travers, M.,  
 Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,  
 Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE  
JOURNAL  
COMMENT

Direct Submission  
 Submitted (22-FEB-2003) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Feb 22, 2003 this sequence version replaced gi:22915576.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIRB  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L16415  
 Center clone name: 105 P 19  
 ----- Summary Statistics  
 Sequencing vector: plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye, 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 205419 bases at least Q40  
 Consensus quality: 208054 bases at least Q30  
 Consensus quality: 208754 bases at least Q20  
 Insert size: 18200; agarose-fp  
 Insert size: 209129; sum-of-coverage  
 Quality coverage: 10.5 in Q20 bases; agarose-fp

Quality coverage: 9.1 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 24 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```
1 20853: contig of 20853 bp in length
* 20854 20953: gap of 100 bp
* 20954 21633: contig of 680 bp in length
* 21634 21733: gap of 100 bp
* 21734 22390: contig of 657 bp in length
* 22391 22491: gap of 100 bp
* 22491 23404: contig of 914 bp in length
* 23405 23505: gap of 100 bp
* 23505 24145: contig of 641 bp in length
* 24146 24245: gap of 100 bp
* 24246 25058: contig of 813 bp in length
* 25059 25158: gap of 100 bp
* 25159 26138: contig of 980 bp in length
* 26139 26238: gap of 100 bp
* 26239 26988: contig of 660 bp in length
* 26989 27865: gap of 100 bp
* 27866 27967: contig of 868 bp in length
* 27968 28590: gap of 100 bp
* 28591 28691: contig of 624 bp in length
* 28692 29543: gap of 100 bp
* 29544 29643: contig of 853 bp in length
* 29644 30707: gap of 1064 bp in length
* 30708 30807: gap of 100 bp
* 30808 32000: contig of 1193 bp in length
* 32001 32101: gap of 100 bp
* 32101 33156: contig of 1056 bp in length
* 33157 33256: gap of 100 bp
* 33257 34255: contig of 1009 bp in length
* 34256 34355: gap of 100 bp
* 34356 35371: contig of 1006 bp in length
* 35372 35472: gap of 100 bp
* 35473 36860: contig of 1389 bp in length
* 36861 36960: gap of 100 bp
* 36961 38773: contig of 1813 bp in length
* 38774 38873: gap of 100 bp
* 38874 40835: contig of 1963 bp in length
* 40836 40936: gap of 100 bp
* 40937 42700: contig of 1764 bp in length
* 42701 42800: gap of 100 bp
* 42801 59066: contig of 16266 bp in length
* 59067 59167: gap of 100 bp
* 59168 94920: contig of 35754 bp in length
* 94921 95020: gap of 100 bp
* 95021 151531: contig of 56511 bp in length
* 151532 151632: gap of 100 bp
* 151633 211429: contig of 59798 bp in length.
```

## FEATURES

## Source

```
1. .211429
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP23-105P19"
/clone_lib="RPCI-23 Female Mouse BAC"
1. .20853
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
20954..21633
/note="assembly_fragment"
21734..22390
/note="assembly_fragment"
22491..23404
misc_feature
```

```
misc_feature /note="assembly_fragment"
23505..24145
/note="assembly_fragment"
misc_feature 24246..25058
/note="assembly_fragment"
misc_feature 25159..26138
/note="assembly_fragment"
misc_feature 26239..26898
/note="assembly_fragment"
misc_feature 26999..27866
/note="assembly_fragment"
misc_feature 27967..28590
/note="assembly_fragment"
misc_feature 28691..29543
/note="assembly_fragment"
misc_feature 29644..30707
/note="assembly_fragment"
misc_feature 30808..32000
/note="assembly_fragment"
misc_feature 32101..33156
/note="assembly_fragment"
misc_feature 33257..34265
/note="assembly_fragment"
misc_feature 34366..35371
/note="assembly_fragment"
misc_feature 35472..36860
/note="assembly_fragment"
misc_feature 36961..38773
/note="assembly_fragment"
misc_feature 38874..40836
/note="assembly_fragment"
misc_feature 40937..42700
/note="assembly_fragment"
misc_feature 42801..59066
/note="assembly_fragment"
misc_feature 59167..94920
/note="assembly_fragment"
misc_feature 95021..151531
/note="assembly_fragment"
misc_feature 151632..211429
/note="assembly_fragment"
```

Query Match 2.6%; Score 57; DB 2; Length 211429;  
Best Local Similarity 100.0%; Pred. No. 1.9e-19;  
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 877 GACCTTTGAGAGCTATCTCTGTCAGACACTGATATACACATTGCTAAAGCTGG 933  
DB 181836 GACCTTTGAGAGCTATCTCTGTCAGACACTGATATACACATTGCTAAAGCTGG 181892

Search completed: July 27, 2004, 16:33:06  
Job time : 8608 secs

***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 27, 2004, 10:49:52 ; Search time 879 Seconds  
(without alignments)  
10748.569 Million cell updates/sec

Title: US-09-714-865B-1  
Perfect score: 2224  
Sequence: 1 actctgaagtcaccatg9999.....aagctctggtttgatgca 2224

Scoring table: OLIGO NUC  
Gapop 60.0 , Gapext 60.0

Searched: 3373863 seqs, 212409041 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_29Jan04.\*

- 1: geneseqn1980s.\*
- 2: geneseqn1990s.\*
- 3: geneseqn2000s.\*
- 4: geneseqn2001as.\*
- 5: geneseqn2001bs.\*
- 6: geneseqn2002s.\*
- 7: geneseqn2003as.\*
- 8: geneseqn2003bs.\*
- 9: geneseqn2003cs.\*
- 10: geneseqn2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2224	100.0	2224	5	AAD06354 Human vas
2	1288	57.9	1984	9	ADB63498 Human CDN
3	579	26.0	1006	3	AAA44726 Human sec
4	393	17.7	677	4	AAS40979 CDNA enco
5	310	13.9	667	7	ABZ20094 Group III
6	278	12.5	591	4	AAS41534 CDNA enco
7	278	12.5	591	4	AAL01169 Human rep
8	278	12.5	591	4	ABL96630 Human tes
9	269	12.1	275	7	ABZ19496 Human tes
10	220	9.9	2194	4	AAL04793 Human rep
11	220	9.9	2194	4	ABL97687 Human tes
12	220	9.9	2195	4	AAL04794 Human rep
13	220	9.9	2195	4	ABL97688 Human tes
14	220	9.9	2197	4	ABL04795 Human rep
15	220	9.9	2197	4	ABL97689 Human tes
16	73	3.3	344	7	ABZ18691 Group III
17	60	2.7	60	6	ABN42617 Human spl
18	36	1.6	1383	7	ACA53957 Prokaryot
19	25	1.1	28	5	AAD06355 Human vas
20	23	1.0	2187	6	AB199243 Mouse isc
21	21	0.9	21	5	AAD06358 RT-PCR pr
22	21	0.9	1842	4	AAS53256 Haemophil
23	21	0.9	1842	7	ACA34056 Prokaryot

24	21	0.9	110000	2	AAT42063_02	Continuation (3 of
25	20	0.9	20	5	AAD06359	Aad06359 RT-PCR pr
26	20	0.9	20	5	AAD06356	Aad06356 PCR prime
27	20	0.9	2320	3	AC37735	Aac37735 Arabidops
28	20	0.9	9832	6	ABL32657	Ab132657 Human Imm
29	20	0.9	96594	9	ADC85257	Adc85257 Mouse Pcp
30	20	0.9	96595	8	ADA02777	Ada02777 Mouse Pcp
31	20	0.9	96595	8	ADB72515	Adb72515 Mouse Pcp
32	20	0.9	110000	6	ABA03041_07	Continuation (8 of
33	19	0.9	28	5	AAD06357	Aad06357 PCR prime
34	19	0.9	138	4	ABA72946	Abat72946 Human toe
35	19	0.9	138	4	AA153370	AA153370 Probe #22
36	19	0.9	138	4	AAK47539	Aak47539 Human Don
37	19	0.9	138	4	AAK21377	Aak21377 Human bra
38	19	0.9	138	4	AB847275	Ab847275 Human liv
39	19	0.9	138	6	AB821638	Ab821638 Human Gen
40	19	0.9	391	7	ACD92737	Acc92737 Human col
41	19	0.9	418	8	ACH48505	Act48505 Human leu
42	19	0.9	504	7	ADA70581	Ada70581 Rice gene
43	19	0.9	592	4	ABA60407	Ab60407 Human foe
44	19	0.9	592	4	AA140291	AA140291 Probe #89
45	19	0.9	592	4	AAK34573	Aak34573 Human bon

ALIGNMENTS

RESULT 1	
AAD06354	
ID	AAD06354 standard; CDNA; 2224 BP.
AC	AAD06354;
XX	
DT	10-AVG-2001 (first entry)
XX	
DE	Human vasa CDNA.
XX	
KW	Human; vasa; germ cell; gonad development; therapy; cancer; oral; brain;
KW	ovarian; biliary tract; breast; pancreas; prostate; colorectal; cervical;
KW	colon; lung; testis; thyroid; oesophageal; endometrial; gastric;
KW	skin; liver; lymphoma; melanoma; glioblastoma; neuroblastoma; neoplasm;
KW	medulloblastoma; choriochorionoma; squamous cell carcinoma; leukemia;
KW	acute lymphocytic; myelogenous; multiple myeloma; Paget's disease;
KW	osteosarcoma; Acquired immune deficiency syndrome; AIDS; Bowen's disease;
KW	leiomyosarcoma; rhabdomyosarcoma; liposarcoma; Kaposi's sarcoma;
KW	fibrosarcoma; tumour; testicular; seminoma; ovarian; dysgerminoma;
XX	teratoma; mediastinal; intracranial; ss.
XX	
OS	Homo sapiens.
XX	
FT	Key
FT	CDS
FT	Location/Qualifiers
FT	/*tag= a
FT	/product= "Human vasa protein"
FT	misc_feature
FT	14..2185
FT	/*tag= b
FT	/note= "This region is specifically claimed as SEQ ID NO:
FT	15 in claim 3 of the specification"
XX	
PN	WO200136445-A1.
XX	
PD	25-MAY-2001.
XX	
PF	16-NOV-2000; 2000WO-US031465.
XX	
PR	18-NOV-1999; 99US-0166394P.
XX	
PA	(BGMN ) BRIGHAM & WOMENS HOSPITAL INC.
XX	
PI	Castrillon DH;
XX	
DR	WPI, 2001-355606/37.
DR	P-PSDB; AAE02417.



XX Novel vasa polynucleotides useful in the diagnosis or treatment of  
PT conditions characterized by aberrant expression and/or presence of mutant  
PT forms of vasa polynucleotides or polypeptides.  
XX  
PS Claim 1; Page 52-53; 66pp; English.  
XX  
CC The present sequence is human vasa cDNA that has germ cell specific  
CC expression and is believed to play a determinative role in gonad  
CC development. Germ cells are specialised to produce haploid gametes in  
CC multicellular organisms. Vasa is useful in the diagnosis or treatment of  
CC conditions characterised by its aberrant expression and/or the presence  
CC of its mutant forms. The conditions include cancers such as biliary  
CC tract, brain, breast, colon, ovarian, pancreas, prostate, colorectal,  
CC oral, liver, lung, skin, basocellular, testis, renal, thyroid, cervical,  
CC endometrial, oesophagel and gastric, lymphomas, melanomas,  
CC glioblastomas, neuroblastomas, medulloblastomas, chorioctarcinoma,  
CC squamous cell carcinoma, haematological neoplasms, acute lymphocytic and  
CC myelogenous leukaemia, multiple myeloma, Acquired immune deficiency  
CC syndrome (AIDS) associated leukemias, interepithelial neoplasms, Bowen's  
CC disease, Paget's disease, sarcomas such as leiomyosarcoma,  
CC rhabdomyosarcoma, liposarcoma, fibrosarcoma, Kaposi's sarcoma and  
CC osteosarcoma, tumours such as testicular tumour (eg. seminoma), ovarian  
CC tumour (eg. dysgerminoma or teratoma) and tumour of an extragonadal  
CC tissue (eg. mediastinal or an intracranial tumour)  
XX  
SQ Sequence 2224 BP; 678 A; 385 C; 550 G; 611 T; 0 U; 0 Other:  
Query Match 100.0%; Score 2224; DB 5; Length 2224;  
Best Local Similarity 100.0%; Pired. No. 0;  
Matches 2224; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 601 TACTTCCAAAGCAAGAGTGGCACTGGAAGTGAACGAGGTGGTTACAAAGTTTAAATGA 660  
QY 661 AGAAGTAATTAACAGGCTCTGGAAAAGAAATCTTGGAACTCAAGACAGAGAGAGAG 720  
DB 661 AGAAGTAATTAACAGGCTCTGGAAAAGAAATCTTGGAACTCAAGAGCAAGAGAGAGAG 720  
QY 721 TAGGTATATCAAGACCAAGCAAAAGTGAACCTTCAATACCCCTCCACCTGAGGTGGA 780  
DB 721 TAGGTATATCAAGACCAAGCAAAAGTGAACCTTCAATACCCCTCCACCTGAGGTGGA 780  
QY 781 CTCATCTTTGACATTAATGACAGGCAATTAATGACAAATATGCACTATTTCTGT 840  
DB 781 CTCATCTTTGACATTAATGACAGGCAATTAATGACAAATATGCACTATTTCTGT 840  
QY 841 GGAAGTGTGGAATATATGACACAGCAATTTGCACTTTTGAAGAATCTATCTGT 900  
DB 841 GGAAGTGTGGAATATATGACACAGCAATTTGCACTTTTGAAGAATCTATCTGT 900  
QY 901 TCACACACTGAATTAACAAATGCTTAAAGTGTATTAATTAATTAATCTTGTGCAAA 960  
DB 901 TCACACACTGAATTAACAAATGCTTAAAGTGTATTAATTAATTAATCTTGTGCAAA 960  
QY 961 ATACAGTATCTTATATCTATCTGACAGACAGATTTGATGCTTGTGCTCAACAGGTC 1020  
DB 961 ATACAGTATCTTATATCTATCTGACAGACAGATTTGATGCTTGTGCTCAACAGGTC 1020  
QY 1021 TGGGAAGACAGGCTTTTCTCTCAATTTTGGCTCATATGATGATGATGATGATGATG 1080  
DB 1021 TGGGAAGACAGGCTTTTCTCTCAATTTTGGCTCATATGATGATGATGATGATGATG 1080  
QY 1081 TGCCAGTGTATTAAGAGTTGACAGAAACAGAGTATTAATTAATTAATTAATTAAT 1140  
DB 1081 TGCCAGTGTATTAAGAGTTGACAGAAACAGAGTATTAATTAATTAATTAATTAAT 1140  
QY 1141 ATTGTCACACAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1200  
DB 1141 ATTGTCACACAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1200  
QY 1201 TGTGTTATATATGAGGGAACCCAGCTGGACATTCATTAATTAATTAATTAATTAAT 1260  
DB 1201 TGTGTTATATATGAGGGAACCCAGCTGGACATTCATTAATTAATTAATTAATTAAT 1260  
QY 1261 TAAATATATATGCTCTCTCTGGAACAGATGATGATGATGATGATGATGATGATGATG 1320  
DB 1261 TAAATATATATGCTCTCTCTGGAACAGATGATGATGATGATGATGATGATGATGATG 1320  
QY 1321 TCTCAACACAGATCAAAATATATGATGATGATGATGATGATGATGATGATGATGATG 1380  
DB 1321 TCTCAACACAGATCAAAATATATGATGATGATGATGATGATGATGATGATGATGATG 1380  
QY 1381 TGTGTCAGAAATGAAGAAATTAATTTTGTGCAAGATGATGATGATGATGATGATGATG 1440  
DB 1381 TGTGTCAGAAATGAAGAAATTAATTTTGTGCAAGATGATGATGATGATGATGATGATG 1440  
QY 1441 AACCTTATGCTCAGTCACTTTTCCAGAGAAATCAAGATGATGATGATGATGATGATG 1500  
DB 1441 AACCTTATGCTCAGTCACTTTTCCAGAGAAATCAAGATGATGATGATGATGATGATG 1500  
QY 1501 AAAGTCAAAATATATGCTTTGCTGTTGGAACAAGTGGGAGCAATGAGAGATGTTCA 1560  
DB 1501 AAAGTCAAAATATATGCTTTGCTGTTGGAACAAGTGGGAGCAATGAGAGATGTTCA 1560  
QY 1561 GCAGACGCTTCCAAATGAGGCACTTTTCCAAAGAGAAAGTGTGTAATTTCTGCG 1620  
DB 1561 GCAGACGCTTCCAAATGAGGCACTTTTCCAAAGAGAAAGTGTGTAATTTCTGCG 1620  
QY 1621 AAACATAGGAGATGAAGAACTATGCTTTGTTGTAACATTAAGAAAGCAATTTTAC 1680  
DB 1621 AAACATAGGAGATGAAGAACTATGCTTTGTTGTAACATTAAGAAAGCAATTTTAC 1680  
QY 1681 TGCAACTTTTCTTGTCAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1740  
DB 1681 TGCAACTTTTCTTGTCAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1740

[illegible]

Pt	Isozaki T, Sugiyama T, Otsuki T, Makamatsu A, Sato H, Ishii S,
Pt	Yanamoto J, Isono Y, Hiro Y, Otsuka K, Nagai K, Itie R, Tamechika I,
Pt	Seki N, Yoshikawa T, Otsuka M, Nagahara K, Masuno Y,
Dx	WPI: 2003-450961/43.
DR	P-PsDB; ADB65468.
XX	
XX	
XX	Claim 1; Page; 222pp; English.
XX	
XX	The invention discloses a polynucleotide comprising a sequence selected
CC	from 1970 fully defined nucleotide sequences which encode novel
CC	polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
CC	or its partial peptide, an antibody binding to the polypeptide or peptide
CC	of the polynucleotide, immunologically assaying the polypeptide or
CC	peptide of the polynucleotide by contacting the polypeptide or peptide
CC	with the antibody, a transformant carrying the polynucleotide in an
CC	expressible manner and an antisense polynucleotide. The oligonucleotide
CC	is useful as a primer for synthesizing the polynucleotide, or as a probe
CC	for detecting the polynucleotide. The polynucleotides and encoded
CC	proteins are useful as pharmaceutical agents and many disease-related
CC	genes may be included in them, for developing a diagnostic marker or
CC	medicines for regulation of their expression and activity, or as targets
CC	of gene therapy. The genes are involved in tissue and/or cell
CC	regeneration. Membrane proteins, signal transduction-related proteins,
CC	transcription-related proteins, disease-related proteins and genes
CC	encoding them can be used as indicators for diseases (e.g. osteoporosis,
CC	neurological diseases, cancer, tumours). The cDNA may be used to regulate
CC	the activity or expression of the encoded protein to treat diseases. The
CC	sequence presented is a cDNA of the invention. Note: Some of the sequence
CC	data for this patent is not represented in the printed specification, but
CC	is based on sequence information supplied by the European Patent Office.
XX	
XQ	Sequence 1984 BP; 612 A; 357 C; 448 G; 567 T; 0 U; 0 Other;
	Query Match 57.9%; Score 1288; DB 9; Length 1984;
	Best Local Similarity 99.7%; Pred. No. 0;
	Matches 1488; Conservative 0; Mismatches 4; Indels 0; Gaps 0
Qy	733 AGAGCAAAAAGTGAACCTACATACCCTCTCCACCTGAGATGAGAGACTCATCTTTGC 792
Dy	292 AGGACCAAAAGTGACTACATACCCCCCTCCTCCACCTGAGATGAGAGACTCATCTTTGC 351
Qy	793 ACATTATCAGACAGGACATAAATCTTCGACAAATAGACACTATTCTTGGAGAAGTCTCG 852
Dy	352 ACATTATCAGACAGGACATAAATCTTCGACAAATAGACACTATTCTTGGAGAAGTCTCG 411
Qy	853 ACATGATGACACACAGCAATTCGTGACTTTTGAAGAAGCTATCTCTGTGCAGACTGAA 912
Dy	412 ACATGATGACACACAGCAATTCGTGACTTTTGAAGAAGCTATCTCTGTGCAGACTGAA 471
Qy	913 TAACAATGCTTAACACTGTTTACTTAAGCTTACTCTGTGCAAAAATACAGTATTTC 972
Dy	472 TAACAATGCTTAACACTGTTTACTTAAGCTTACTCTGTGCAAAAATACAGTATTTC 531
Qy	973 TATCATACTGACAGAGAGATTTGATGGCTTGTGCTCAACAGAGTCTGGGAAGACTGC 1032
Dy	532 TATCATACTGACAGAGAGATTTGATGGCTTGTGCTCAACAGAGTCTGGGAAGACTGC 591
Qy	1033 GGCTTTCTCTCCATCAATTTTGGCTATATGATGCAATGAGATTAATCTGCAGTCGTTT 1093
Dy	592 GGCTTTCTCTCCATCAATTTTGGCTATATGATGCAATGAGATTAATCTGCAGTCGTTT 651
Qy	1093 TAAAGAGTTGCAGAACACAGAGTATATATTTGAGACCAACTGAGAAATGGTCAACCA 1153
Dy	652 TAAAGAGTTGCAGAACACAGAGTATATATTTGAGACCAACTGAGAAATGGTCAACCA 711
Qy	1153 GATTATTATGAGGCCGAATTTCTTTTGGACTGTGCTAGAGCTGTTGTTATATA 1212

Db 712 GATTATTTGGAGAACGAGAAATTTCTTTGGGACTGTGTAGAGCTGTGTATATA 771

Qy 1213 TGGGGAGAACGAGCTGGGACATTCATTCGACAAATAGTACAGGCTGTATATATATG 1272

Db 772 TGGGGAGAACGAGCTGGGACATTCATTCGACAAATAGTACAGGCTGTATATATATG 831

Qy 1273 TGTCTACTCTGGAGAGCTGATGATATCTATAGGCAAGAAAGATTTGGTCTCAACAGAT 1332

Db 832 TGTCTACTCTGGAGAGCTGATGATATCTATAGGCAAGAAAGATTTGGTCTCAACAGAT 891

Qy 1333 CAATATCTTATGTTTGGATGAGCTGATGATGATGATGATGATGATGATGATGATGAT 1392

Db 892 CAATATCTTATGTTTGGATGAGCTGATGATGATGATGATGATGATGATGATGATGAT 951

Qy 1393 GAGAGATTAATTTCTTGGCCAGAGATTCGATCAAGAGAGAGAGAGAGAGAGAGAGAGAT 1452

Db 952 GAGAGATTAATTTCTTGGCCAGAGATTCGATCAAGAGAGAGAGAGAGAGAGAGAGAGAT 1011

Qy 1453 CAGTGGACCTTTCCAGAGAGAGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1512

Db 1012 CAGTGGACCTTTCCAGAGAGAGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1071

Qy 1513 TCTGTTTGTCTGTTGGAGCAAGTGGGTGGAGCATGTAGAGATTTCAAGAGAGAGATTTCT 1572

Db 1072 TCTGTTTGTCTGTTGGAGCAAGTGGGTGGAGCATGTAGAGATTTCAAGAGAGAGATTTCT 1131

Qy 1573 CCAAGTTGGCCAGTCTCTCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1632

Db 1132 CCAAGTTGGCCAGTCTCTCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1191

Qy 1633 TGAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1692

Db 1192 TGAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1251

Qy 1693 TTGTCAAG 1752

Db 1252 TTGTCAAG 1311

Qy 1753 GCAAGCTCTTGGAGATTTGCTTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1812

Db 1312 GCAAGCTCTTGGAGATTTGCTTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1371

Qy 1813 TGGCAG 1872

Db 1372 TGGCAG 1431

Qy 1873 CATTTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1932

Db 1432 CATTTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1491

Qy 1933 AATTTCTTTTGTGATCTTGAATCGATTAACATTTAGCAGAGCTCTAGTAAAGTATTT 1992

Db 1492 AATTTCTTTTGTGATCTTGAATCGATTAACATTTAGCAGAGCTCTAGTAAAGTATTT 1551

Qy 1993 GACAGATGCTCAACAGAGATTTCTCTGATGATGATGATGATGATGATGATGATGATGAT 2052

Db 1552 GACAGATGCTCAACAGAGATTTCTCTGATGATGATGATGATGATGATGATGATGATGAT 1611

Qy 2053 TCTGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2112

Db 1612 TCTGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1671

Qy 2113 CAGAGAGATTTGAACAGAT 2172

Db 1672 CAGAGAGATTTGAACAGAT 1731

Qy 2173 TGAATCATGGAGATTAAGAT 2234

Db 1732 TGAATCATGGAGATTAAGAT 1793

ID AAA44726 standard; cDNA; 1006 BP.

XX AAA44726;

AC 21-AUG-2000 (first entry)

DT 21-AUG-2000 (first entry)

XX Human secreted expressed sequence tag SEQ ID NO:1301.

DE Human; mouse; chicken; rat; secreted expressed sequence tag; SEST; expressed sequence tag; SEST; probe; chemotactic; proliferative; immunomodulatory; haematopoietic; chemokine; analgesic; haemostatic; chemokine; anti-inflammatory; cytostatic; antibacterial; antifungal; antiviral; antidiabetic; antiaspartic; vulnary; antiparkinsonian; antitumor; osteoprotective; neuroprotective; nocotropic; antipsoriatic; cerebroprotective; anticonvulsant; antidepressant; gene therapy; vaccine; autoimmune disorder; multiple sclerosis; allergic condition; insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer; lymphoid cell deficiency; burn; osteoporosis; osteoarthritis; central nervous system disorder; Alzheimer's disease; stroke; Parkinson's disease; Huntington's disease; coagulation disorder; haemophilia; thrombosis; inflammatory disorder; Crohn's disease; tumour; infection; depression; psoriasis; ss.

XX Homo sapiens.

OS WO20021991-A1.

PN 20-APR-2000.

XX 15-OCT-1999; 99WO-US024206.

XX 15-OCT-1999; 98US-010436P.

PR 15-OCT-1998; 98US-010436P.

XX (GENY) GENETICS INST INC.

PA Jacobo K, McCoy JM, Lavalie ER, Collins-Racie LA, Evans C, Meiberg D, Treacy M, Bowman MR,

PI MPI, 2000-317938/27.

DR Isolated polynucleotides, and encoded proteins, comprising secreted expressed sequence tags (SESTs), useful for treating various disorders such as autoimmune, infectious, and central nervous system disorders.

PT Claim 1; Page 540; 803pp; English.

XX AAA44726 to AAA45925 represent specifically claimed secreted expressed sequence tags (SESTs), isolated from human, mouse, chicken and rat tissue sources. The SESTs can have a range of activities depending on the tissues they were isolated from. The activities include: chemotactic; proliferative; immunomodulatory; haematopoietic; chemokine; analgesic; haemostatic; thrombolytic; anti-inflammatory; cytostatic; antibacterial; antifungal; antiviral; antidiabetic; antiaspartic; vulnary; antitumor; osteoprotective; neuroprotective; nocotropic; antiparkinsonian; antipsoriatic; cerebroprotective; anticonvulsant; and antidepressant. The SESTs can be used for gene therapy and in vaccines. The SESTs are useful as probes for the identification and isolation of full-length cDNAs and genomic DNA molecules which correspond to the SESTs. Proteins encoded by the SESTs are useful in assays for determining biological activity and raising antibodies. They may be useful for treatment of autoimmune disorders (multiple sclerosis, insulin dependent diabetes), allergic conditions (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers, osteoporosis, osteoarthritis, central nervous system disorders (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's disease), tumours, bacterial, fungal or viral infections, depression and psoriasis. AAA45926 to AAA45931 represent linker variants which are given in the exemplification of the present invention

CC Sequence 1006 BP; 314 A; 181 C; 251 G; 260 T; 0 U; 0 Other;

XX Query Match 26.0%; Score 579; DB 3; Length 1006;

XX Best Local Similarity 99.8%; Pred. No. 5.3e-276;

Wed Jul 28 11:42:29 2004

us-09-714-865b-1.rng

Page 5

Matches 629; Conservative 0; Mismatches 1; Indels 0; Gaps 0;				
Qy	510	ATAATGACTTAGACCGACGATGATGACGACGACGCTTTTGGTTCTGTAA	569	
Db	377	ATAATGACTTAGACCGACGATGATGACGACGACGCTTTTGGTTCTGTAA	436	
Qy	570	GACCAATTAATGAGGACAGGATAGGTGATCTTCTCAAGAGAGAGTGGACGTGAA	629	
Db	437	GACCAATTAATGAGGACAGGATAGGTGATCTTCTCAAGAGAGAGTGGACGTGAA	496	
Qy	630	GTGAACGAGGTGTTCAAGTTTAATGAGAGATTAATACGGCTCTGGAAGAAAT	689	
Db	497	GTGAACGAGGTGTTCAAGTTTAATGAGAGATTAATACGGCTCTGGAAGAAAT	556	
Qy	690	CTTGAAGTCAAGAGCAGAGAGAGAGAAAGTACTGATCTCAAGACCAAAAGTACT	749	
Db	557	CTTGAAGTCAAGAGCAGAGAGAGAGAAAGTACTGATCTCAAGACCAAAAGTACT	616	
Qy	750	ACATACCCCTCCCTCCACCTGAGATGAGACTCTTTCACATTATCAGACGCA	809	
Db	617	ACATACCCCTCCCTCCACCTGAGATGAGACTCTTTCACATTATCAGACGCA	676	
Qy	810	TAACTTGCACAAATACGACACTATCTTGTGAAAGTGTCTGACATGATGACAC	869	
Db	677	TAACTTGCACAAATACGACACTATCTTGTGAAAGTGTCTGACATGATGACAC	736	
Qy	870	CAATCTGACTTTTGAAGAACTATCTCTGTACGACACTGATTAACACATTGCTAA	929	
Db	737	CAATCTGACTTTTGAAGAACTATCTCTGTACGACACTGATTAACACATTGCTAA	796	
Qy	930	CTGGTTACTTAAGCTTACTCCGTGTGCAAAATACATATCTTCCATATCTTGGAC	989	
Db	797	CTGGTTACTTAAGCTTACTCCGTGTGCAAAATACATATCTTCCATATCTTGGAC	856	
Qy	990	GAGATTGATGCTTTGCTCAACAGAGCTGTGGAAGACTGGGCTTTTCTCTCA	1049	
Db	857	GAGATTGATGCTTTGCTCAACAGAGCTGTGGAAGACTGGGCTTTTCTCTCA	916	
Qy	1050	TTTTGGCTCATATGATGATGATGATTAATCTGCACTGCTTTTAAAGTTGAC	1109	
Db	917	TTTTGGCTCATATGATGATGATGATTAATCTGCACTGCTTTTAAAGTTGAC	976	
Qy	1110	CAGAGTATATATGATGACCAACTGAG	1139	
Db	977	CAGAGTATATATGATGACCAACTGAG	1006	

PR	04-FEB-2000;	2000US-0180628P.
PR	24-FEB-2000;	2000US-0184664P.
PR	02-MAR-2000;	2000US-0186350P.
PR	16-MAR-2000;	2000US-0189874P.
PR	17-MAR-2000;	2000US-0190076P.
PR	18-APR-2000;	2000US-0198123P.
PR	19-MAY-2000;	2000US-0205155P.
PR	07-JUN-2000;	2000US-0209467P.
PR	28-JUN-2000;	2000US-0214886P.
PR	30-JUN-2000;	2000US-0215135P.
PR	07-JUL-2000;	2000US-0216647P.
PR	07-JUL-2000;	2000US-0216880P.
PR	11-JUL-2000;	2000US-0217487P.
PR	11-JUL-2000;	2000US-0217966P.
PR	14-JUL-2000;	2000US-0218290P.
PR	26-JUL-2000;	2000US-0220963P.
PR	26-JUL-2000;	2000US-0220964P.
PR	14-AUG-2000;	2000US-0224518P.
PR	14-AUG-2000;	2000US-0224519P.
PR	14-AUG-2000;	2000US-0225213P.
PR	14-AUG-2000;	2000US-0225214P.
PR	14-AUG-2000;	2000US-0225266P.
PR	14-AUG-2000;	2000US-0225267P.
PR	14-AUG-2000;	2000US-0225268P.
PR	14-AUG-2000;	2000US-0225270P.
PR	14-AUG-2000;	2000US-0225477P.
PR	14-AUG-2000;	2000US-0225757P.
PR	14-AUG-2000;	2000US-0225758P.
PR	14-AUG-2000;	2000US-0225759P.
PR	18-AUG-2000;	2000US-0226279P.
PR	22-AUG-2000;	2000US-0226681P.
PR	22-AUG-2000;	2000US-0226686P.
PR	23-AUG-2000;	2000US-0227182P.
PR	23-AUG-2000;	2000US-0227009P.
PR	30-AUG-2000;	2000US-0228924P.
PR	01-SEP-2000;	2000US-0229287P.
PR	01-SEP-2000;	2000US-0229343P.
PR	01-SEP-2000;	2000US-0229344P.
PR	01-SEP-2000;	2000US-0229345P.
PR	05-SEP-2000;	2000US-0229509P.
PR	05-SEP-2000;	2000US-0229513P.
PR	06-SEP-2000;	2000US-0230437P.
PR	06-SEP-2000;	2000US-0230438P.
PR	08-SEP-2000;	2000US-0231243P.
PR	08-SEP-2000;	2000US-0231244P.
PR	08-SEP-2000;	2000US-0231413P.
PR	08-SEP-2000;	2000US-0231414P.
PR	08-SEP-2000;	2000US-0232080P.
PR	08-SEP-2000;	2000US-0232081P.
PR	12-SEP-2000;	2000US-0231968P.
PR	14-SEP-2000;	2000US-0232397P.
PR	14-SEP-2000;	2000US-0232398P.
PR	14-SEP-2000;	2000US-0232399P.
PR	14-SEP-2000;	2000US-0232400P.
PR	14-SEP-2000;	2000US-0232401P.
PR	14-SEP-2000;	2000US-0232402P.
PR	14-SEP-2000;	2000US-0233064P.
PR	14-SEP-2000;	2000US-0233065P.
PR	21-SEP-2000;	2000US-0234223P.
PR	21-SEP-2000;	2000US-0234274P.
PR	25-SEP-2000;	2000US-0234997P.
PR	25-SEP-2000;	2000US-0234998P.
PR	26-SEP-2000;	2000US-0234984P.
PR	27-SEP-2000;	2000US-0235834P.
PR	27-SEP-2000;	2000US-0235835P.
PR	29-SEP-2000;	2000US-0236327P.
PR	29-SEP-2000;	2000US-0236367P.
PR	29-SEP-2000;	2000US-0236368P.
PR	29-SEP-2000;	2000US-0236369P.
PR	29-SEP-2000;	2000US-0236370P.
PR	02-OCT-2000;	2000US-0236802P.
PR	02-OCT-2000;	2000US-0236803P.

RESULT 4  
AAS40979 ID AAS40979 standard; cDNA; 677 BP.  
XX AAS40979;  
AC  
DT 17-DEC-2001 (first entry)  
XX  
XX cDNA encoding novel human enzyme polypeptide #195.  
DE  
XX  
XX Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;  
KM ligase; hyperproliferative disorder; immunodeficiency disorder;  
KM autoimmune disorder; neurological disorder; metabolic disorder;  
KM inflammatory disorder; cardiovascular disorder; reproductive disorder;  
KM blood-related disorder; infectious disorder; gene therapy; cytostatic;  
KM anti arthritic; nephrotropic; anticoagulant; ss.  
XX  
XX Homo sapiens.  
OS  
XX  
XX W0200155301-A2.  
PN  
XX 02-AUG-2001.  
PD  
XX  
XX 17-JAN-2001; 2001WO-US001239.  
PF  
XX  
XX 31-JAN-2000; 2000US-0179065P  
PR



KM recombinant DNA technology; gene; ss.  
XX  
XX Homo sapiens.  
OS  
XX WO200278516-A2.  
FN  
XX 10-OCT-2002.  
PD  
XX  
XX 28-MAR-2002; 2002WO-US010421.  
PF  
XX 30-MAR-2001; 2001US-0280255P.  
PR 28-AUG-2001; 2001US-0315563P.  
PR 09-JAN-2002; 2002US-0347313P.  
XX  
XX (CORI-) CORIXA CORP.  
PA  
XX Wang T, Wang S, Bangur CS, Gaiger A;  
PI  
XX WPI; 2003-058387/05.  
DR  
XX  
XX New immunogenic polynucleotides or polypeptides useful for diagnosing, in  
PT preventing and treating cancer expressing CT or CP mRNA antigens, and, in  
PT virology, immunology, microbiology, molecular biology and recombinant DNA  
PT techniques.  
XX  
XX Claim 1; SEQ ID NO 2520; 207pp; English.  
XX  
XX AB017575 to AB020506 represent isolated polynucleotide (I) sequences, and  
CC AB054446 to AB054472 represent protein (II) sequences, from the present  
CC invention. (I) and (II) have cytostatic activity and can be used in gene  
CC therapy and vaccines. (I), (II), antibodies and compositions from the  
CC present invention are useful for diagnosing, preventing and treating  
CC cancer, which expresses CT or CP mRNA antigens. They are useful for  
CC stimulating immune response. They can also be useful in virology,  
CC immunology, microbiology, molecular biology and recombinant DNA  
CC techniques. N.B. The sequence data for this patent did not form part of  
CC the printed specification, but was obtained in electronic format directly  
CC from WIPO at [http://wipo.int/pub/published\\_pcl\\_sequences](http://wipo.int/pub/published_pcl_sequences)  
XX  
SQ Sequence 667 BP; 215 A; 146 C; 155 G; 150 T; 0 U; 1 Other;  
Query Match 13.9%; Score 310; DB 7; Length 667;  
Best Local Similarity 100.0%; Pred. No. 1e-142;  
Matches 310; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 655 AAATGAAGAGTAATACAGGCTCTGGAAGAATTTCTGAGTCAAGACAGAGAGG 714  
Db 1 AAATGAAGAGTAATACAGGCTCTGGAAGAATTTCTGAGTCAAGACAGAGAGG 60  
QY 715 AGAAGTAGTATCTCAAGACCAAGAGTACTATACCCCTCTCTCACTTAGGA 714  
Db 61 AGAAGTAGTATCTCAAGACCAAGAGTACTATACCCCTCTCTCACTTAGGA 120  
QY 775 TGAGGATCTCACTTTGCACTTATGACAGCAGCAATTCGACAAATACGACTAT 834  
Db 121 TGAGGATCTCACTTTGCACTTATGACAGCAGCAATTCGACAAATACGACTAT 180  
QY 835 TCTTGTGGAAGTGTCTGACATGATGACACCAAGCAATTCGACCTTTGAAGACTAA 894  
Db 181 TCTTGTGGAAGTGTCTGACATGATGACACCAAGCAATTCGACCTTTGAAGACTAA 240  
QY 895 TCTGTGACAGCACTGATTAACAACATTGCTAAAGCTGTTATCTAGTACTCTCTGT 954  
Db 241 TCTGTGACAGCACTGATTAACAACATTGCTAAAGCTGTTATCTAGTACTCTCTGT 300  
QY 955 GCAAAATATAC 964  
Db 301 GCAAAATATAC 310

RESULT 6  
AAS41534  
ID AAS41534 standard; cDNA; 591 BP.

XX  
AC AAS41534;  
XX  
DT 17-DEC-2001 (first entry)  
XX  
DE cDNA encoding novel human enzyme polypeptide #750.  
XX  
XX Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;  
XX ligase; hyperproliferative disorder; immunodeficiency disorder;  
XX autoimmune disorder; neurological disorder; metabolic disorder;  
XX inflammatory disorder; cardiovascular disorder; reproductive disorder;  
XX blood-related disorder; infectious disorder; gene therapy; cytostatic;  
XX anti arthritic; nephrotropic; anticoagulant; ss.  
OS  
XX Homo sapiens.  
XX  
XX WO200155301-A2.  
PN  
XX  
XX 02-AUG-2001.  
PD  
XX  
XX 17-JAN-2001; 2001WO-US001239.  
PF  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-MAR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-020515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214866P.  
PR 30-JUL-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216809P.  
PR 11-UTL-2000; 2000US-0217487P.  
PR 11-UTL-2000; 2000US-0217496P.  
PR 14-UTL-2000; 2000US-0218290P.  
PR 26-UTL-2000; 2000US-0220963P.  
PR 26-UTL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225477P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226686P.  
PR 22-AUG-2000; 2000US-0227183P.  
PR 23-AUG-2000; 2000US-0227095P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.

```

PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234224P.
PR 21-SEP-2000; 2000US-0234977P.
PR 25-SEP-2000; 2000US-0234988P.
PR 25-SEP-2000; 2000US-0234989P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246529P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249247P.
PR 17-NOV-2000; 2000US-0249249P.
PR 17-NOV-2000; 2000US-0249299P.
PR 01-DEC-2000; 2000US-0250160P.

```

```

PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 05-DEC-2000; 2000US-0251479P.
PR 06-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254997P.
PR 03-JAN-2001; 2001US-0259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI, 2001-465566/50.
XX P-PSDB; AAU23664.
XX
XX Novel polypeptides and polynucleotides useful for diagnosing, preventing,
XX treating neural, immune system, muscular, reproductive, pulmonary,
XX cardiovascular, renal, proliferative disorders and cancerous diseases.
XX
XX Claim 4; SEQ ID NO 760; 1180pp; English.
XX
XX The present invention relates to the isolation of novel human enzyme
XX polypeptides (AAU22915-AAU23614), and the cDNA and genomic sequences
XX encoding them. The enzyme polypeptides of the invention may comprise the
XX functional classes of oxidoreductases, transferases, hydrolases, lyases,
XX isomerases or ligases. The sequences of the invention are useful in the
XX diagnosis, treatment, prevention and/or prognosis of a wide range of
XX disorders including hyperproliferative disorders (e.g. cancer),
XX immunodeficiency disorders (e.g. AIDS), autoimmune disorders (e.g.
XX arthritis), neurological disorders (e.g. Alzheimer's disease), metabolic
XX disorders (e.g. phenylketonuria), inflammatory disorders (e.g. asthma),
XX cardiovascular disorders (e.g. atherosclerosis), blood-related disorders
XX (e.g. haemophilia), reproductive disorders (e.g. infertility) and
XX infectious disorders (e.g. influenza). The polynucleotides of the
XX invention can also be used in gene therapy. AA54785-AA54184 represent
XX cDNA sequences encoding for the novel human enzyme polypeptides of the
XX invention. Note: The sequence data for this patent did not form part of
XX the printed specification, but was obtained in electronic format directly
XX from WIP0 at ftp.wip0.int/pub/published_pcl_sequences
XX
XX Sequence 591 BP; 203 A; 100 C; 111 G; 176 T; 0 U; 1 Other;
XX
XX Query Match 12.5%; Score 278; DB 4; Length 591;
XX Best Local Similarity 99.7%; Pred. No. 7.2e-127;
XX Matches 328; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 1896 TTGGGCGTACCTGGCTGTTGGGGAATCTGCGAGCAATTTCTTTTATCTTGAAT 1955
XX 30 TTGGGCGTACCTGGCTGTTGGGGAATCTGCGAGCAATTTCTTTTATCTTGAAT 89
XX
XX 1956 CGGATAACCATTTAGACAGAGCTCTAGTAAGATTTAGACAGATGCTCAAGAGATGTC 2015
XX 90 CGGATAACCATTTAGACAGAGCTCTAGTAAGATTTAGACAGATGCTCAAGAGATGTC 149
XX
XX 2016 CTGATGTTGGAGAAATGCTTGTATACATATCTTGTCTCAGTGGTAGTACAA 2075
XX 150 CTGATGTTGGAGAAATGCTTGTATACATATCTTGTCTCAGTGGTAGTACAA 209
XX
XX 2076 GAGGAAAGCTTTGATCATGATATACAGAAAGGCAAGACACTTTGAAACAGCTG 2135
XX 210 GAGGAAAGCTTTGATCATGATATACAGAAAGGCAAGACACTTTGAAACAGCTG 269
XX
XX 2136 GGTTCCTTCTTACAGAGCTCCCAATCATAGATATGATGATGATTAAGCCAAA 2195
XX 270 GGTTCCTTCTTACAGAGCTCCCAATCATAGATATGATGATGATTAAGCCAAA 329
XX
XX ACATCCTCAAGTCTGCTTGTATGCA 2224
XX

```



Db 330 ACATCCTTCAGTCTGCGTTTGATGCA 358

RESULT 7  
AAL01169  
ID AAL01169 standard; cDNA; 591 BP.  
XX  
AC AAL01169;  
XX  
DT 21-NOV-2001 (first entry)  
XX  
DE Human reproductive system related antigen cDNA SHQ ID NO: 1170.  
XX  
KW Human; reproductive system related antigen; reproductive system disorder;  
KW cancer; gene therapy; ss.  
XX  
OS Homo sapiens.  
XX  
PN MO20015330-A2.  
XX  
PD 02-AUG-2001.  
XX  
FF 17-JAN-2001; 2001WO-US001339.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225477P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226568P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0232402P.  
PR 14-SEP-2000; 2000US-0232403P.  
PR 14-SEP-2000; 2000US-0232404P.  
PR 14-SEP-2000; 2000US-0232405P.  
PR 14-SEP-2000; 2000US-0232406P.  
PR 14-SEP-2000; 2000US-0232407P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0234984P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235835P.  
PR 29-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0235837P.  
PR 29-SEP-2000; 2000US-0235838P.  
PR 29-SEP-2000; 2000US-0235839P.  
PR 29-SEP-2000; 2000US-0235840P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239933P.  
PR 13-OCT-2000; 2000US-0239934P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0244674P.  
PR 08-NOV-2000; 2000US-0244675P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249219P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249246P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.

17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251088P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 06-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-465570/50.  
DR P-PSDB; AAM55199.  
XX  
XX Isolated nucleic acid molecule encoding a reproductive system antigen is  
PT used in preventing, treating or ameliorating a medical condition.  
XX  
PS Claim 1; SEQ ID NO 1170; 1297bp + Sequence Listing; English.  
XX  
XX The present invention provides the protein and coding sequences of a  
CC number of human reproductive system related antigens. These can be used  
CC in the prevention and treatment of reproductive system disorders,  
CC including cancer. The present sequence is a coding sequence of the  
CC invention  
XX  
SQ Sequence 591 BP; 203 A; 100 C; 111 G; 176 T; 0 U; 1 Other;  
Query Match 12.5%; Score 278; DB 4; Length 591;  
Best Local Similarity 99.7%; Pred. No. 7,2e-127;  
Matches 328; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1896 TTGGGCGTACTGCTGTTGGGGAATCTGGCAGACCAATTCTTTTGTCTTGAAT 1955  
DB 30 TTGGGCGTACTGCTGTTGGGGAATCTGGCAGACCAATTCTTTTGTCTTGAAT 89  
QY 1956 CGGATACCATTTAGACAGACCCCTCTAGTAAAGTATTGACAGTGTCTCAACGATGTTTC 2015  
DB 90 CGGATACCATTTAGACAGACCCCTCTAGTAAAGTATTGACAGTGTCTCAACGATGTTTC 149  
QY 2016 CTGCATGTTGGAAGAAATTCCTTTAGTACATACATTCCTGCTTCAGTGTAGTACAA 2075  
DB 150 CTGCATGTTGGAAGAAATTCCTTTAGTACATACATTCCTGCTTCAGTGTAGTACAA 209  
QY 2076 GAGGAAACGTTGTCATGTCATGATCCAGAAAGGCAAGGACCTTTGAACACAGCTG 2135  
DB 210 GAGGAAACGTTGTCATGTCATGATCCAGAAAGGCAAGGACCTTTGAACACAGCTG 269  
QY 2136 GGTTCCTTCTTCACGAGCTCCCAATCCAGTAGATGATGATGATTAAGCCAAA 2195  
DB 270 GGTTCCTTCTTCACGAGCTCCCAATCCAGTAGATGATGATGATTAAGCCAAA 329  
QY 2196 ACAATCTTCAAGTCTGTGTTTGAATGCA 2224  
DB 330 ACAATCTTCAAGTCTGTGTTTGAATGCA 358  
RESULT 8  
ABL96630  
ID ABL96630 standard; cDNA; 591 BP.  
XX  
XX ABL96630;  
XX  
XX 21-JUN-2002 (first entry)  
XX  
DE Human testicular antigen encoding cDNA SEQ ID NO: 298.

XX  
XX Human; testicular antigen; testes; cancer; metastasis; immune disorder;  
XX reproductive system disorder; urinary system disorder; gene therapy;  
XX cardiovascular disorder; respiratory disorder; neurological disorder;  
XX gastrointestinal disease; infection; cytostatic; gene; ss.  
OS Homo sapiens.  
XX  
XX WO20015317-A2.  
XX  
XX 02-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US001329.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
XX 04-FEB-2000; 2000US-0180628P.  
XX 24-FEB-2000; 2000US-0184664P.  
XX 02-MAR-2000; 2000US-0186350P.  
XX 16-MAR-2000; 2000US-0189874P.  
XX 17-MAR-2000; 2000US-0190076P.  
XX 18-APR-2000; 2000US-0198133P.  
XX 19-MAY-2000; 2000US-0205515P.  
XX 07-JUN-2000; 2000US-0209467P.  
XX 28-JUN-2000; 2000US-0214886P.  
XX 30-JUN-2000; 2000US-0215135P.  
XX 07-JUL-2000; 2000US-0216647P.  
XX 07-JUL-2000; 2000US-0216880P.  
XX 11-JUL-2000; 2000US-0217487P.  
XX 11-JUL-2000; 2000US-0217496P.  
XX 14-JUL-2000; 2000US-0218290P.  
XX 26-JUL-2000; 2000US-0220963P.  
XX 26-JUL-2000; 2000US-0220964P.  
XX 14-AUG-2000; 2000US-0224518P.  
XX 14-AUG-2000; 2000US-0224519P.  
XX 14-AUG-2000; 2000US-0225213P.  
XX 14-AUG-2000; 2000US-0225214P.  
XX 14-AUG-2000; 2000US-0225266P.  
XX 14-AUG-2000; 2000US-0225267P.  
XX 14-AUG-2000; 2000US-0225268P.  
XX 14-AUG-2000; 2000US-0225270P.  
XX 14-AUG-2000; 2000US-0225447P.  
XX 14-AUG-2000; 2000US-0225757P.  
XX 14-AUG-2000; 2000US-0225758P.  
XX 14-AUG-2000; 2000US-0225759P.  
XX 18-AUG-2000; 2000US-0226279P.  
XX 22-AUG-2000; 2000US-0226681P.  
XX 22-AUG-2000; 2000US-0226686P.  
XX 22-AUG-2000; 2000US-0227182P.  
XX 23-AUG-2000; 2000US-0227009P.  
XX 30-AUG-2000; 2000US-0228242P.  
XX 01-SEP-2000; 2000US-0229287P.  
XX 01-SEP-2000; 2000US-0229343P.  
XX 01-SEP-2000; 2000US-0229344P.  
XX 01-SEP-2000; 2000US-0229345P.  
XX 05-SEP-2000; 2000US-0229509P.  
XX 05-SEP-2000; 2000US-0229513P.  
XX 06-SEP-2000; 2000US-0230437P.  
XX 06-SEP-2000; 2000US-0230438P.  
XX 08-SEP-2000; 2000US-0231242P.  
XX 08-SEP-2000; 2000US-0231243P.  
XX 08-SEP-2000; 2000US-0231244P.  
XX 08-SEP-2000; 2000US-0231413P.  
XX 08-SEP-2000; 2000US-0231414P.  
XX 08-SEP-2000; 2000US-0232080P.  
XX 08-SEP-2000; 2000US-0232081P.  
XX 12-SEP-2000; 2000US-0231968P.  
XX 14-SEP-2000; 2000US-0232397P.  
XX 14-SEP-2000; 2000US-0232398P.  
XX 14-SEP-2000; 2000US-0232399P.  
XX 14-SEP-2000; 2000US-0232400P.  
XX 14-SEP-2000; 2000US-0232401P.  
XX 14-SEP-2000; 2000US-0233063P.  
XX 14-SEP-2000; 2000US-0233064P.



PD 10-OCT-2002.  
XX  
XX 28-MAR-2002; 2002WO-US010421.  
XX  
XX 30-MAR-2001; 2001US-0280255P.  
XX 28-AUG-2001; 2001US-0315563P.  
XX 09-JAN-2002; 2002US-0347313P.  
XX  
XX (CORI-) CORIXA CORP.  
XX  
XX Wang T, Wang S, Bangur CS, Gaiger A,  
XX WPI; 2003-058387/05.  
XX  
XX New immunogenic polynucleotides or polypeptides useful for diagnosing,  
XX PT preventing and treating cancer expressing CT or CP mRNA antigens, and in  
XX PT virology, immunology, microbiology, molecular biology and recombinant DNA  
XX PT techniques.  
XX  
XX Claim 1; SEQ ID NO 1922; 207pp; English.  
XX  
XX ABQ17575 to ABQ20506 represent isolated polynucleotide (I) sequences, and  
XX CC ABP54446 to ABP54472 represent protein (II) sequences, from the present  
XX CC invention. (I) and (II) have cytostatic activity and can be used in gene  
XX CC therapy and vaccines. (I), (II), antibodies and compositions from the  
XX CC present invention are useful for diagnosing, preventing and treating  
XX CC cancer, which expresses CT or CP mRNA antigens. They are useful for  
XX CC stimulating immune response. They can also be useful in virology,  
XX CC immunology, microbiology, molecular biology and recombinant DNA  
XX CC techniques. N.B. The sequence data for this patent did not form part of  
XX CC the printed specification, but was obtained in electronic format directly  
XX CC from WIPO at [http://wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences)  
XX  
XX Sequence 275 BP; 87 A; 46 C; 64 G; 78 T; 0 U; 0 Other;  
SQ  
Query Match 12.1%; Score 269; DB 7; Length 275;  
Best Local Similarity 100.0%; Pred. No. 2.1e-122;  
Matches 269; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1232 CATTCAATTGCAATAGTACAGGCTGTAATATTTATGCTCTCTCTGGAAGACTG 1291  
Db 7 CATTCAATTGCAATAGTACAGGCTGTAATATTTATGCTCTCTCTGGAAGACTG 66  
QY 1292 ATGGATTCATAGGCAAGAAAGATTGCTCAACAGATCAAACTAGTTTGGAT 1351  
Db 67 ATGGATTCATAGGCAAGAAAGATTGCTCAACAGATCAAACTAGTTTGGAT 126  
QY 1352 GAAGCTGATGCATGTTGATGATGGTTTGGTCCAGAAATGAAAGATTATTTCTTGC 1411  
Db 127 GAAGCTGATGCATGTTGATGATGGTTTGGTCCAGAAATGAAAGATTATTTCTTGC 186  
QY 1412 CCGGGAATGCCATCAAGAAAGACGCCCAACCTATGTTGAGTCAATTTCCAGAG 1471  
Db 187 CCGGGAATGCCATCAAGAAAGACGCCCAACCTATGTTGAGTCAATTTCCAGAG 246  
QY 1472 GAATTCMAAGTTGGCTGCAGAGTTT 1500  
Db 247 GAATTCMAAGTTGGCTGCAGAGTTT 275  
RESULT 10  
AAL04793  
ID AAL04793 standard; DNA; 2194 BP.  
XX  
XX AAL04793;  
XX  
XX 21-NOV-2001 (first entry)  
XX  
XX Human reproductive system related antigen DNA SEQ ID NO: 7481.  
XX  
XX Human reproductive system related antigen; reproductive system disorder;  
KW cancer; gene therapy; ds.  
XX

OS Homo sapiens.  
XX  
XX WO200155320-A2.  
XX  
XX 02-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US001339.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
XX 04-FEB-2000; 2000US-0180628P.  
XX 24-FEB-2000; 2000US-0184664P.  
XX 02-MAR-2000; 2000US-0186350P.  
XX 16-MAR-2000; 2000US-0189874P.  
XX 17-MAR-2000; 2000US-0190076P.  
XX 18-APR-2000; 2000US-0198133P.  
XX 19-MAY-2000; 2000US-0205151P.  
XX 07-JUN-2000; 2000US-0209467P.  
XX 28-JUN-2000; 2000US-0214886P.  
XX 30-JUN-2000; 2000US-0215135P.  
XX 07-JUL-2000; 2000US-0216647P.  
XX 07-JUL-2000; 2000US-0216880P.  
XX 11-JUL-2000; 2000US-0217487P.  
XX 11-JUL-2000; 2000US-0217496P.  
XX 14-JUL-2000; 2000US-0218290P.  
XX 26-JUL-2000; 2000US-0220963P.  
XX 26-JUL-2000; 2000US-0220964P.  
XX 14-AUG-2000; 2000US-0224518P.  
XX 14-AUG-2000; 2000US-0224519P.  
XX 14-AUG-2000; 2000US-0225213P.  
XX 14-AUG-2000; 2000US-0225214P.  
XX 14-AUG-2000; 2000US-0225266P.  
XX 14-AUG-2000; 2000US-0225267P.  
XX 14-AUG-2000; 2000US-0225268P.  
XX 14-AUG-2000; 2000US-0225270P.  
XX 14-AUG-2000; 2000US-0225447P.  
XX 14-AUG-2000; 2000US-0225757P.  
XX 14-AUG-2000; 2000US-0225758P.  
XX 14-AUG-2000; 2000US-0225759P.  
XX 18-AUG-2000; 2000US-0226279P.  
XX 22-AUG-2000; 2000US-0226681P.  
XX 22-AUG-2000; 2000US-0226868P.  
XX 22-AUG-2000; 2000US-0227182P.  
XX 23-AUG-2000; 2000US-0227009P.  
XX 30-AUG-2000; 2000US-0228924P.  
XX 01-SEP-2000; 2000US-0229287P.  
XX 01-SEP-2000; 2000US-0229343P.  
XX 01-SEP-2000; 2000US-0229344P.  
XX 01-SEP-2000; 2000US-0229345P.  
XX 05-SEP-2000; 2000US-0229509P.  
XX 05-SEP-2000; 2000US-0229513P.  
XX 06-SEP-2000; 2000US-0230437P.  
XX 06-SEP-2000; 2000US-0230438P.  
XX 08-SEP-2000; 2000US-0231242P.  
XX 08-SEP-2000; 2000US-0231243P.  
XX 08-SEP-2000; 2000US-0231244P.  
XX 08-SEP-2000; 2000US-0231413P.  
XX 08-SEP-2000; 2000US-0231414P.  
XX 08-SEP-2000; 2000US-0232080P.  
XX 08-SEP-2000; 2000US-0232081P.  
XX 12-SEP-2000; 2000US-0231968P.  
XX 14-SEP-2000; 2000US-0232397P.  
XX 14-SEP-2000; 2000US-0232398P.  
XX 14-SEP-2000; 2000US-0232399P.  
XX 14-SEP-2000; 2000US-0232400P.  
XX 14-SEP-2000; 2000US-0232401P.  
XX 14-SEP-2000; 2000US-0233063P.  
XX 14-SEP-2000; 2000US-0233064P.  
XX 14-SEP-2000; 2000US-0233065P.  
XX 21-SEP-2000; 2000US-0234223P.  
XX 21-SEP-2000; 2000US-0234274P.  
XX 25-SEP-2000; 2000US-0234997P.  
XX 25-SEP-2000; 2000US-0234998P.  
XX 26-SEP-2000; 2000US-0235484P.

PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0235327P.  
PR 29-SEP-2000; 2000US-0235367P.  
PR 29-SEP-2000; 2000US-0235368P.  
PR 29-SEP-2000; 2000US-0235369P.  
PR 29-SEP-2000; 2000US-0235370P.  
PR 29-SEP-2000; 2000US-0235370P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239337P.  
PR 13-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241825P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0244747P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249298P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251866P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2000US-0255676P.  
XX  
XX  
PR (HUMA-) HUMAN GENOME SCT INC.

XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-465570/50.  
DR  
XX  
XX  
PT Isolated nucleic acid molecule encoding a reproductive system antigen is  
PT used in preventing, treating or ameliorating a medical condition.  
XX  
XX  
PS Disclosure; SEQ ID NO 7481; 1297bp + Sequence Listing; English.  
XX  
XX  
CC The present invention provides the protein and coding sequences of a  
CC number of human reproductive system related antigens. These can be used  
CC in the prevention and treatment of reproductive system disorders,  
CC including cancer. The present sequence is a genomic sequence encoding a  
CC protein of the invention  
XX  
SQ Sequence 2194 BP, 624 A, 349 C, 400 G, 821 T; 0 U; 0 Other;  
  
Query Match 9.9%; Score 220; DB 4; Length 2194;  
Best Local Similarity 100.0%; Pred. No. 3.9e-98;  
Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1781 AAGTGGCCAGTCTTTTCTGCTACTTACGTAAGCTGCCAGGAGGCTGGATTTGAAATGCG 1840  
DB 1 AAGTGGCCAGTCTTTTCTGCTACTTACGTAAGCTGCCAGGAGGCTGGATTTGAAATGCG 60  
  
QY 1841 CAACATGTTATCAATTTTGTATCTTCTTACCATTTGATGAATGATGCAATTTGGG 1900  
DB 61 CAACATGTTATCAATTTTGTATCTTCTTACCATTTGATGAATGATGCAATTTGGG 120  
  
QY 1901 CGTACTGTCGTGTGGGAATCTGACAGAGCAATTTCTTTTGTATCTTGAATCGAT 1960  
DB 121 CGTACTGTCGTGTGGGAATCTGACAGAGCAATTTCTTTTGTATCTTGAATCGAT 180  
  
QY 1961 AACCATTTAGCAGAGCTCTAGTAAAGATTTAGCAGATG 2000  
DB 181 AACCATTTAGCAGAGCTCTAGTAAAGATTTAGCAGATG 220  
  
RESULT 11  
ABJ97687  
ID ABL97687 standard; DNA; 2194 BP.  
AC  
XX ABL97687;  
AC  
XX  
DT 21-JUN-2002 (first entry)  
XX  
XX  
DE Human testicular antigen encoding DNA fragment SEQ ID NO: 2339.  
XX  
XX  
KW Human; testicular antigen; testes; cancer; metastasis; immune disorder;  
KW reproductive system disorder; urinary system disorder; gene therapy;  
KW cardiovascular disorder; respiratory disorder; neurological disorder;  
XX Gastrointestinal disease; infection; cytostatic; gene; ds.  
OS Homo sapiens.  
XX  
PN WO200155317-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US001329.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-APR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.



Query Match 9.9%; Score 220; DB 4; Length 2194;  
Best Local Similarity 100.0%; Pred. No. 3,9e-98;  
Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1781 AAGTCCAGTCTTGTGTCTACTTCACTGCTGCGAGGGCTGGATATTGAAATGTG 1840  
DB 1 AAGTCCAGTCTTGTGTCTACTTCACTGCTGCGAGGGCTGGATATTGAAATGTG 60  
QY 1841 CAACATTTATCAATTTGATCTTCTCTACATGATGATATGTTCAATCGAATTGGG 1900  
DB 61 CAACATTTATCAATTTGATCTTCTCTACATGATGATATGTTCAATCGAATTGGG 120  
QY 1901 CGTACTGTCGTCGTGGGAATCTGGACAGAGCAATTCCTTTTGTGATCTTGATTCGGAT 1960  
DB 121 CGTACTGTCGTCGTGGGAATCTGGACAGAGCAATTCCTTTTGTGATCTTGATTCGGAT 180  
QY 1961 AACCATTTAGCAGAGCCTCTAGTAAAGTATTGACAGATG 2000  
DB 181 AACCATTTAGCAGAGCCTCTAGTAAAGTATTGACAGATG 220

RESULT 12

AA04794  
ID AA04794 standard; DNA: 2195 BP.

XX AA04794;  
DT 21-NOV-2001 (first entry)

XX Human reproductive system related antigen DNA SEQ ID NO: 7482.

XX Human; reproductive system related antigen; reproductive system disorder;  
KM cancer; gene therapy; ds.

XX Homo sapiens.

XX W0200155320-A2.

XX PD 02-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US001339.

XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217486P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.

PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226688P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239335P.  
PR 13-OCT-2000; 2000US-0239337P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246603P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.





PR 08-SEP-2000; 2000US-0233080P.  
PR 08-SEP-2000; 2000US-0233081P.  
PR 12-SEP-2000; 2000US-02331968P.  
PR 14-SEP-2000; 2000US-0233397P.  
PR 14-SEP-2000; 2000US-0233398P.  
PR 14-SEP-2000; 2000US-0233399P.  
PR 14-SEP-2000; 2000US-0233400P.  
PR 14-SEP-2000; 2000US-0233401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239355P.  
PR 13-OCT-2000; 2000US-0239357P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.

PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251888P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251889P.  
PR 08-DEC-2000; 2000US-0251900P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
  
(HUMA-) HUMAN GENOME SCI INC.  
XX  
XX  
PI Rosen CA, Barash SC, Ruben SW,  
XX  
XX WPI; 2001-483232/52.  
XX  
XX Nucleic acids encoding 973 human testicular antigen polypeptides, useful  
PT for preventing, diagnosing and/or treating testicular cancer.  
XX  
XX PS Disclosure; SEQ ID NO 2340; 766pp; English.  
XX  
XX CC The present invention provides the protein and coding sequences of 973  
CC human testicular antigens, and fragments of their genomic sequences. The  
CC sequences can be used in the treatment of cardiovascular, urinary system,  
CC reproductive system, immune, respiratory, neurological and  
CC gastrointestinal disorders, infections, and particularly a cancer,  
CC especially testicular cancers. The present sequence is a DNA encoding a  
CC protein fragment of the invention  
XX  
SQ Sequence 2195 BP; 624 A; 350 C; 400 G; 821 T; 0 U; 0 Other;  
  
Query Match 9.9%; Score 220; DB 4; Length 2195;  
Best Local Similarity 100.0%; Pred. No. 3.9e-98;  
Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1781 AAGTCCAGTCTTGTGCTACTTCACTAGTGCAGAGGCTGATTAATGAATGTG 1840  
DB 1 AAGTCCAGTCTTGTGCTACTTCACTAGTGCAGAGGCTGATTAATGAATGTG 60  
  
QY 1841 CAACATGTTATCAATTTTGATTTCTTCTTCACTAGTGAATATGTTCAATTTGGG 1900  
DB 61 CAACATGTTATCAATTTTGATTTCTTCTTCACTAGTGAATATGTTCAATTTGGG 120  
  
QY 1901 CGTACTGTCGTTGTGGAACTAGTGCAGAGCAATTTCTTTTGATTCGATTCGAT 1960  
DB 121 CGTACTGTCGTTGTGGAACTAGTGCAGAGCAATTTCTTTTGATTCGATTCGAT 180  
  
QY 1961 AACCATTTAGCACAGCTCTTGTAGTAAAGTATTCAGATG 2000  
DB 181 AACCATTTAGCACAGCTCTTGTAGTAAAGTATTCAGATG 220  
  
RESULT 14  
ID AAL04795 standard; DNA; 2197 BP.  
XX  
XX AAL04795;  
AC  
XX  
DT 21-NOV-2001 (first entry)  
XX  
XX  
DE Human reproductive system related antigen DNA SEQ ID NO: 7463.  
XX  
XX Human; reproductive system related antigen; reproductive system disorder;  
KW cancer; gene therapy; db.  
XX  
OS Homo sapiens.  
XX  
XX W020015320-A2.

XX 02-AUG-2001.  
PD 17-JAN-2001; 2001WO-US001339.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-018464P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0190123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214866P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226688P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 05-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 06-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.

PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 01-NOV-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249246P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249266P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0251989P.  
PR 06-DEC-2000; 2000US-0251719P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251858P.  
PR 08-DEC-2000; 2000US-0251863P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
PA (HUMA-) HUMAN GENOME SCI. INC.  
XX  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX

DR WPI; 2001-465570/50.  
XX Isolated nucleic acid molecule encoding a reproductive system antigen is  
PT used in preventing, treating or ameliorating a medical condition.  
XX  
XX Disclosure; SEQ ID NO 7483; 1297bp + Sequence Listing; English.  
XX  
XX The present invention provides the protein and coding sequences of a  
CC number of human reproductive system related antigens. These can be used  
CC in the prevention and treatment of reproductive system disorders,  
CC including cancer. The present sequence is a genomic sequence encoding a  
CC protein of the invention  
XX  
SQ Sequence 2197 BP; 624 A; 349 C; 401 G; 823 T; 0 U; 0 Other;  
  
Query Match 9.9%; Score 220; DB 4; Length 2197;  
Best Local Similarity 100.0%; Pred. No. 3,9e-98;  
Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1781 AAGTCCAGTTCTGTGCTACTTCACTAGTACCTGCGCAGAGGCGGATTTGAAATGTG 1840  
DB 1 AAGTGCCAGTTCTGTGCTACTTCACTAGTACCTGCGCAGAGGCGGATTTGAAATGTG 60  
  
QY 1841 CAACATGTTATCAATTTTGATCTTCCTTCAACATGATGAAATGTTCAATGCGG 1900  
DB 61 CAACATGTTATCAATTTTGATCTTCCTTCAACATGATGAAATGTTCAATGCGG 120  
  
QY 1901 CGTCTGGTGGTGTGGGAATACGCGAGAGCAATTCCTTTTGTATCTTGATCGGAT 1960  
DB 121 CGTCTGTGCTGTGTGGAAATCTGCGAGCAATTTCCCTTTTGTATCTTGATCGGAT 180  
  
QY 1961 AACCATTTAGCACAGCCTCTAGTAAAGTATTGACAGATG 2000  
DB 181 AACCATTTAGCACAGCCTCTAGTAAAGTATTGACAGATG 220  
  
RESULT 15  
ABL97689  
ID ABL97689 standard; DNA; 2197 BP.  
XX  
XX ABL97689;  
AC  
XX  
XX 21-JUN-2002 (first entry)  
DT  
XX  
XX Human testicular antigen encoding DNA fragment SEQ ID NO: 2341.  
DE  
XX  
XX Human; testicular antigen; testes; cancer; metastasis; immune disorder;  
KM reproductive system disorder; urinary system disorder; gene therapy;  
KW cardiovascular disorder; respiratory disorder; neurological disorder;  
XX gastrointestinal disease; infection; cytoskeletal; gene; ds.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200155317-A2.  
PN  
XX  
XX 02-AUG-2001.  
PD  
XX  
XX 17-JAN-2001; 2001WO-US001329.  
PF  
XX  
XX  
31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.

PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225477P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226682P.  
PR 23-AUG-2000; 2000US-0227182P.  
PR 30-AUG-2000; 2000US-0227709P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0233397P.  
PR 14-SEP-2000; 2000US-0233398P.  
PR 14-SEP-2000; 2000US-0233399P.  
PR 14-SEP-2000; 2000US-0234002P.  
PR 14-SEP-2000; 2000US-023401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-023497P.  
PR 25-SEP-2000; 2000US-023498P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235835P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.

PR 01-NOV-2000; 2000US-0244617P.  
 PR 08-NOV-2000; 2000US-0246474P.  
 PR 08-NOV-2000; 2000US-0246475P.  
 PR 08-NOV-2000; 2000US-0246476P.  
 PR 08-NOV-2000; 2000US-0246477P.  
 PR 08-NOV-2000; 2000US-0246478P.  
 PR 08-NOV-2000; 2000US-0246479P.  
 PR 08-NOV-2000; 2000US-0246523P.  
 PR 08-NOV-2000; 2000US-0246524P.  
 PR 08-NOV-2000; 2000US-0246525P.  
 PR 08-NOV-2000; 2000US-0246526P.  
 PR 08-NOV-2000; 2000US-0246527P.  
 PR 08-NOV-2000; 2000US-0246528P.  
 PR 08-NOV-2000; 2000US-0246532P.  
 PR 08-NOV-2000; 2000US-0246609P.  
 PR 08-NOV-2000; 2000US-0246610P.  
 PR 08-NOV-2000; 2000US-0246611P.  
 PR 08-NOV-2000; 2000US-0246613P.  
 PR 17-NOV-2000; 2000US-0249207P.  
 PR 17-NOV-2000; 2000US-0249208P.  
 PR 17-NOV-2000; 2000US-0249209P.  
 PR 17-NOV-2000; 2000US-0249210P.  
 PR 17-NOV-2000; 2000US-0249211P.  
 PR 17-NOV-2000; 2000US-0249212P.  
 PR 17-NOV-2000; 2000US-0249213P.  
 PR 17-NOV-2000; 2000US-0249214P.  
 PR 17-NOV-2000; 2000US-0249215P.  
 PR 17-NOV-2000; 2000US-0249216P.  
 PR 17-NOV-2000; 2000US-0249217P.  
 PR 17-NOV-2000; 2000US-0249218P.  
 PR 17-NOV-2000; 2000US-0249244P.  
 PR 17-NOV-2000; 2000US-0249245P.  
 PR 17-NOV-2000; 2000US-0249246P.  
 PR 17-NOV-2000; 2000US-0249265P.  
 PR 17-NOV-2000; 2000US-0249297P.  
 PR 17-NOV-2000; 2000US-0249299P.  
 PR 17-NOV-2000; 2000US-0249300P.  
 PR 01-DEC-2000; 2000US-0250160P.  
 PR 01-DEC-2000; 2000US-0250391P.  
 PR 05-DEC-2000; 2000US-0251030P.  
 PR 05-DEC-2000; 2000US-0251988P.  
 PR 05-DEC-2000; 2000US-0256719P.  
 PR 06-DEC-2000; 2000US-0251479P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251868P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 PR 08-DEC-2000; 2000US-0251899P.  
 PR 08-DEC-2000; 2000US-0251990P.  
 PR 11-DEC-2000; 2000US-0254097P.  
 PR 05-JAN-2001; 2001US-0259678P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX  
 DR WPI; 2001-483332/52.  
 XX  
 PT Nucleic acids encoding 973 human testicular antigen polypeptides, useful  
 PT for preventing, diagnosing and/or treating testicular cancer.  
 XX  
 PS Disclosure; SEQ ID NO 2341; 766bp; English.  
 XX  
 CC The present invention provides the protein and coding sequences of 973  
 CC human testicular antigens, and fragments of their genomic sequences. The  
 CC sequences can be used in the treatment of cardiovascular, urinary system,  
 CC reproductive system, immune, respiratory, neurological and  
 CC gastrointestinal disorders, infections, and particularly cancer,  
 CC especially testicular cancers. The present sequence is a DNA encoding a  
 CC protein fragment of the invention  
 XX  
 SQ Sequence 2197 BP; 624 A; 349 C; 401 G; 823 T; 0 U; 0 Other;

Query March 9.9%; Score 220; DB 4; Length 2197;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-98;

Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1781 AAGTCCAGTCTTGTGCTACTTCACTGAGTGCAGAGGCTGGATATTTGAAAATGTG 1840  
 |||||  
 Db 1 AAGTCCAGTCTTGTGCTACTTCACTGAGTGCAGAGGCTGGATATTTGAAAATGTG 60  
 QY 1841 CAACATGTATTCATTTTGTATCTTCTTCACTGATGTAATATGTCATCGAATTTGG 1900  
 |||||  
 Db 61 CAACATGTATTCATTTTGTATCTTCTTCACTGATGTAATATGTCATCGAATTTGG 120  
 QY 1901 CGTACTGATGCTGTGGGAATACGTGACAGACATTTCTTTTGTATCTTGAATCGAT 1960  
 |||||  
 Db 121 CGTACTGATGCTGTGGGAATACGTGACAGACATTTCTTTTGTATCTTGAATCGAT 180  
 QY 1961 AACCATTTAGCACAGCCTCTAGTAAAGTATTTGACAGATG 2000  
 |||||  
 Db 181 AACCATTTAGCACAGCCTCTAGTAAAGTATTTGACAGATG 220

Search completed: July 27, 2004, 14:09:31  
 Job time : 885 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 27, 2004, 13:19:58 ; Search time 166 Seconds  
(without alignments)  
7435.005 Million cell updates/sec

Title: US-09-714-865B-1  
Perfect score: 2224  
Sequence: 1 acttgagtcacccatg9959.....aagctctgtgttgcata 2224

Scoring table: OLIGO NUC  
Gapop 60.0 , Gapext 60.0

Searched: 682709 seqs, 277475446 residues

Word size: 0

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

- Database :
- 1: Issued Patents NA.\*
  - 2: /cgn2\_6/ptodata/2/ina/5A COMB.seq.\*
  - 3: /cgn2\_6/ptodata/2/ina/5B COMB.seq.\*
  - 4: /cgn2\_6/ptodata/2/ina/6A COMB.seq.\*
  - 5: /cgn2\_6/ptodata/2/ina/6B COMB.seq.\*
  - 6: /cgn2\_6/ptodata/2/ina/PCTUS.COMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	448	20.1	504	US-09-621-976-2923	Sequence 2923, Ap
2	340	15.3	420	US-08-833-381-1475	Sequence 1475, Ap
3	21	0.9	1830121	US-09-557-884-1	Sequence 1, Appl
4	21	0.9	1830121	US-09-643-990A-1	Sequence 1, Appl
5	19	0.9	2319	US-09-058-489-90	Sequence 90, Appl
6	19	0.9	2451	US-09-976-594-786	Sequence 786, Ap
7	19	0.9	2850	US-09-184-964-2	Sequence 2, Appl
8	19	0.9	3408	US-09-058-489-14	Sequence 14, Appl
9	19	0.9	4416	US-09-058-489-17	Sequence 17, Appl
10	19	0.9	5322	US-09-058-489-13	Sequence 13, Appl
11	19	0.9	536165	US-09-214-808-1	Sequence 1, Appl
12	18	0.8	339	US-08-651-155B-116	Sequence 116, App
13	18	0.8	339	US-09-194-036B-116	Sequence 116, App
14	18	0.8	421	US-09-232-575-116	Sequence 116, App
15	18	0.8	421	US-09-389-681-116	Sequence 116, App
16	18	0.8	421	US-09-620-405B-116	Sequence 116, App
17	18	0.8	421	US-09-339-338-116	Sequence 116, App
18	18	0.8	421	US-09-433-826B-116	Sequence 116, App
19	18	0.8	421	US-09-604-287A-116	Sequence 116, App
20	18	0.8	421	US-09-285-480-116	Sequence 116, App
21	18	0.8	421	US-09-834-759-116	Sequence 116, App
22	18	0.8	532	US-09-318-443-3	Sequence 3, Appl
23	18	0.8	598	US-09-833-381-99	Sequence 99, Appl
24	18	0.8	813	US-09-540-236-226	Sequence 226, App
25	18	0.8	1205	US-09-533-029-5	Sequence 5, Appl
26	18	0.8	1536	US-09-318-443-5	Sequence 5, Appl
27	18	0.8	1682	US-09-318-443-7	Sequence 7, Appl

C	28	18	0.8	1796	3	US-09-255-392-1	Sequence 1, Appl
	29	18	0.8	1876	3	US-09-082-092-5	Sequence 5, Appl
	30	18	0.8	1876	4	US-09-885-722A-5	Sequence 5, Appl
	31	18	0.8	1914	4	US-09-291-922-19	Sequence 19, Appl
C	32	18	0.8	2005	2	US-08-484-200-1	Sequence 1, Appl
C	33	18	0.8	2005	2	US-08-477-493-1	Sequence 1, Appl
C	34	18	0.8	2005	3	US-08-465-375-1	Sequence 1, Appl
C	35	18	0.8	2005	3	US-08-465-375-1	Sequence 1, Appl
C	36	18	0.8	2005	4	US-10-142-373-1	Sequence 1, Appl
C	37	18	0.8	2478	4	US-09-328-352-1848	Sequence 1848, Ap
C	38	18	0.8	3001	4	US-09-539-333D-151	Sequence 151, App
	39	18	0.8	3103	3	US-08-826-246-3	Sequence 3, Appl
	40	18	0.8	3103	3	US-08-944-495-3	Sequence 3, Appl
	41	18	0.8	3103	3	US-09-126-640-2	Sequence 2, Appl
	42	18	0.8	3103	3	US-08-925-588-3	Sequence 3, Appl
	43	18	0.8	3103	4	US-09-288-292A-2	Sequence 2, Appl
	44	18	0.8	3103	4	US-09-372-044-3	Sequence 3, Appl
	45	18	0.8	3103	4	US-08-825-466-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1  
US-09-621-976-2923  
; Sequence 2923 Application US/09621976  
; Patent No. 6639063  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Maline Edwards, J.B.  
; APPLICANT: Gioberto, S.  
; APPLICANT: Gioberto, S.  
; TITLE OF INVENTION: ESTE and Encoded Human Proteins.  
; FILE REFERENCE: GENSET 054PR2  
; CURRENT APPLICATION NUMBER: US/09/621,976  
; NUMBER OF SEQ ID NOS: 2000-07-21  
; SOFTWARE: Patent.pm  
; SEQ ID NO 2923  
; LENGTH: 504  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; NAME/KEY: CDS  
; LOCATION: 224..472  
US-09-621-976-2923

Query Match	20.1%	Score 448	DB 4	Length 504
Best Local Similarity	100.0%	Pred. No. 2.9e-216		
Matches 448	Conservative	0	Mismatches	0
			Indels	0
			Gaps	0
QY	775	TTAGGACTCCATCTTGACATTTATCAGACGAGGTAAGTTGACAAATGACACATAT	834	
DB	1	TGAGACTCCATCTTGACATTTATCAGACGAGGTAAGTTGACAAATGACACATAT	60	
QY	835	TTTGTGAGAGTGTCTGACATGATGACACCAACGAAATTTGACTTTGAGAGCTTA	894	
DB	61	TTTGTGAGAGTGTCTGACATGATGACACCAACGAAATTTGACTTTGAGAGCTTA	120	
QY	895	TCTGTGAGACGATGATGACATGATGACACCAACGAAATTTGACTTTGAGAGCTTA	954	
DB	121	TTTGTGAGACGATGATGACATGATGACACCAACGAAATTTGACTTTGAGAGCTTA	180	
QY	955	GCAAAATGAGATTTCTATCATCTTGACAGACGAGATTTGAGCTTGCTGCTCAAC	1014	
DB	181	GCAAAATGAGATTTCTATCATCTTGACAGACGAGATTTGAGCTTGCTGCTCAAC	240	
QY	1015	AGGCTGAGAGAGAGTGGGCTTTTCTTCAATTTGGCTCATATGATGATGATG	1074	
DB	241	AGGCTGAGAGAGAGTGGGCTTTTCTTCAATTTGGCTCATATGATGATGATG	300	
QY	1075	AATAGCTCCAGTCTTTTAAAGAGTTGACAGAGACGAGATTTATTTAGACACAC	1134	
DB	301	AATAGCTCCAGTCTTTTAAAGAGTTGACAGAGACGAGATTTATTTAGACACAC	360	

QY 1135 TCAGAGATTGTCACACAGATTATTGGAAGCCAGAAATTTCTTTGGACTTGTC 1194  
Db 361 TCAGAGATTGTCACACAGATTATTGGAAGCCAGAAATTTCTTTGGACTTGTC 420  
QY 1195 AAGAGCTGTTGTTATATATGGGGGAACC 1222  
Db 421 AAGAGCTGTTGTTATATATGGGGGAACC 448

RESULT 2  
US-09-833-381-1475  
Sequence 1475, Application US/09833381  
Patent No. 6672186  
GENERAL INFORMATION:  
APPLICANT: Robison, Keith E.  
TITLE OF INVENTION: No. 6672186el Nucleic Acid and Protein Homologs  
FILE REFERENCE: 5800-119  
CURRENT APPLICATION NUMBER: US/09/833,381  
CURRENT FILING DATE: 2001-04-11  
PRIOR APPLICATION NUMBER: 09/516,448  
PRIOR FILING DATE: 2000-02-29  
NUMBER OF SEQ ID NOS: 2050  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 1475  
LENGTH: 420  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-833-381-1475

Query Match 15.3%; Score 340; DB 4; Length 420;  
Best Local Similarity 99.7%; Pred. No. 1,1e-161;  
Matches 390; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 820 CAAATACGACATATTCTTGTGGAAGTCTGGAATGATGACCCACCAATTTCTGAC 879  
Db 30 CAAATACGACATATTCTTGTGGAAGTCTGGAATGATGACCCACCAATTTCTGAC 89  
QY 880 TTTTGAAGAGCTTAATCTCTGACACACTGATTAACAACATTTGCTAAAGCTGTTATAC 939  
Db 90 TTTTGAAGAGCTTAATCTCTGACACACTGATTAACAACATTTGCTAAAGCTGTTATAC 149  
QY 940 TAAGCTTACTCTGTCGCAAAATATCAATATCTTATCAATCTTGAAGAGATTTGAT 999  
Db 150 TAAGCTTACTCTGTCGCAAAATATCAATATCTTATCAATCTTGAAGAGATTTGAT 209  
QY 1000 GCGTTGCTCAAAACAGGCTCTGGAAGACTGCGGCTTTCTCTCAATTTGGCTCA 1059  
Db 210 GCGTTGCTCAAAACAGGCTCTGGAAGACTGCGGCTTTCTCTCAATTTGGCTCA 269  
QY 1060 TATGATCATGATGATGATTAATCTCCAGTCTTTTAAAGAGTTCAGAGAACCAAGTGTAT 1119  
Db 270 TATGATCATGATGATGATTAATCTCCAGTCTTTTAAAGAGTTCAGAGAACCAAGTGTAT 329  
QY 1120 TATTGTAGCAACCACTGCAATTTGTCACACCAATTTTGGAGCCAGAAATTTTC 1179  
Db 330 TATTGTAGCAACCACTGCAATTTGTCACACCAATTTTGGAGCCAGAAATTTTC 389  
QY 1180 TTTTGGAGCTGTTGTAAGAGCTGTTGTATA 1210  
Db 390 TTTTGGAGCTGTTGTAAGAGCTGTTGTATA 420

RESULT 3  
US-09-557-884-1  
Sequence 1, Application US/09557884  
Patent No. 6506581  
GENERAL INFORMATION:  
APPLICANT: Fleischmann et al.  
TITLE OF INVENTION: The Nucleotide sequence of the Haemophilus influenzae Rd Genome, Fragments Thereof, and Uses Thereof  
NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: MD  
COUNTRY: USA  
ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: 3 1/2 inch diskette  
COMPUTER: Dell Pentium  
OPERATING SYSTEM: MS DOS v6.22  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/557,884  
FILING DATE: 25-Apr-2000  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/476,102  
FILING DATE: JUN-5-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Michelle S. Marks  
REGISTRATION NUMBER: 41,971  
REFERENCE/DOCKET NUMBER: PB186P3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 301-309-8504  
TELEFAX: 301-309-8439

SEQUENCE CHARACTERISTICS:

LENGTH: 1830121 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-557-884-1

Query Match 0.9%; Score 21; DB 4; Length 1830121;  
Best Local Similarity 100.0%; Pred. No. 2.4;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1892 CGAATGGGCGTACTGCTGT 1912  
Db 261968 CGAATGGGCGTACTGCTGT 261988

RESULT 4

US-09-643-990A-1  
Sequence 1, Application US/09643990A  
Patent No. 6528289

GENERAL INFORMATION:

APPLICANT: Robert D. Fleischmann  
Mark D. Adams  
Owen White  
Hamilton O. Smith  
J. Craig Venter

TITLE OF INVENTION: The Nucleotide sequence of the Haemophilus influenzae Rd Genome, Fragments Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville,  
STATE: MD  
COUNTRY: USA  
ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: 3 1/2 inch diskette  
COMPUTER: Dell Pentium  
OPERATING SYSTEM: MS DOS v6.22  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/643,990A  
FILING DATE: 23-Aug-2000



```

;
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,429
; FILING DATE: 1995-06-07
; APPLICATION NUMBER: 08/426,787
; FILING DATE: 1995-04-21
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB186P1C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-610-5790
; TELEFAX: 310-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

Query Match          0.9%: Score 21; DB 4; Length 1830121;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1892 CGAATTGGCGTACTGTCGT 1912
DB      261968 CGAATTGGCGTACTGTCGT 261988

RESULT 5
US-09-058-489-90
; Sequence 90, Application US/09058489
; Patent No. 6103886
; GENERAL INFORMATION:
; APPLICANT: Whitehead Institute for Biomedical Research
; APPLICANT: Page, David
; TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
; FILE REFERENCE: WHI97-089A
; CURRENT APPLICATION NUMBER: US/09/058,489
; CURRENT FILING DATE: 1998-04-10
; EARLIER APPLICATION NUMBER: 60/041,877
; EARLIER FILING DATE: 1997-04-11
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 90
; LENGTH: 2319
; TYPE: DNA
; ORGANISM: Human
US-09-058-489-90

Query Match          0.9%: Score 19; DB 3; Length 2319;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1843 ACAATGTTATCAATTTGAT 1861
DB      1598 ACAATGTTATCAATTTGAT 1616

RESULT 6
US-09-976-594-786
; Sequence 786, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
```

```

;
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 786
; LENGTH: 2451
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. 6673549 035282CB1
US-09-976-594-786

Query Match          0.9%: Score 19; DB 4; Length 2451;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1843 ACAATGTTATCAATTTGAT 1861
DB      1572 ACAATGTTATCAATTTGAT 1590

RESULT 7
US-09-184-964-2
; Sequence 2, Application US/09184964
; Patent No. 6391574
; GENERAL INFORMATION:
; APPLICANT: Rine, Jasper D.
; APPLICANT: Boyatchuk, Victor L.
; APPLICANT: Ashby, Matthew N.
; TITLE OF INVENTION: APC1 AND RCE1: ISOPRENYLATED CAXX
; TITLE OF INVENTION: PROCESSING ENZYMES
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Fish & Richardson P.C.
; STREET: 2200 Sand Hill road, suite 100
; CITY: Menlo Park
; STATE: CA
; COUNTRY: USA
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/184,964
; FILING DATE: 03-NOV-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/902,774
; FILING DATE: 30-JUL-1997
; APPLICATION NUMBER: 60/023,491
; FILING DATE: 07-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Syat, Reginald J.
; REGISTRATION NUMBER: 28,172
; REFERENCE/DOCKET NUMBER: 09272-006004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/322-5070
; TELEFAX: 650/854-0875
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2850 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: Genomic DNA
US-09-184-964-2

Query Match          0.9%: Score 19; DB 4; Length 2850;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 1687 TTTCTTTTGTCAAGAAAA 1705  
Db 919 TTTCTTTTGTCAAGAAAA 937

## RESULT 8

US-09-058-489-14  
; Sequence 14, Application US/09058489  
; Patent No. 610386  
; GENERAL INFORMATION:  
; APPLICANT: Whitehead Institute for Biomedical Research  
; APPLICANT: Lahn, Bruce  
; APPLICANT: Page, David  
; TITLE OF INVENTION: Genes in the No. 610386-Recombining Region of  
; FILE REFERENCE: WH197-08PA  
; CURRENT APPLICATION NUMBER: US/09/058,489  
; CURRENT FILING DATE: 1998-04-10  
; EARLIER APPLICATION NUMBER: 60/041,877  
; EARLIER FILING DATE: 1997-04-11  
; NUMBER OF SEQ ID NOS: 91  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 14  
; LENGTH: 3408  
; TYPE: DNA  
; ORGANISM: Human  
US-09-058-489-14

Query Match 0.9%; Score 19; DB 3; Length 3408;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1842 AACATGTTATCAATTTTGA 1860  
Db 2388 AACATGTTATCAATTTTGA 2406

## RESULT 9

US-09-058-489-17  
; Sequence 17, Application US/09058489  
; Patent No. 610386  
; GENERAL INFORMATION:  
; APPLICANT: Whitehead Institute for Biomedical Research  
; APPLICANT: Lahn, Bruce  
; APPLICANT: Page, David  
; TITLE OF INVENTION: Genes in the No. 610386-Recombining Region of  
; FILE REFERENCE: WH197-08PA  
; CURRENT APPLICATION NUMBER: US/09/058,489  
; CURRENT FILING DATE: 1998-04-10  
; EARLIER APPLICATION NUMBER: 60/041,877  
; EARLIER FILING DATE: 1997-04-11  
; NUMBER OF SEQ ID NOS: 91  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 17  
; LENGTH: 4416  
; TYPE: DNA  
; ORGANISM: Human  
US-09-058-489-17

Query Match 0.9%; Score 19; DB 3; Length 4416;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1843 ACATGTTATCAATTTTGA 1861  
Db 1598 ACATGTTATCAATTTTGA 1616

RESULT 10  
US-09-058-489-13  
; Sequence 13, Application US/09058489

; Patent No. 610386  
; GENERAL INFORMATION:  
; APPLICANT: Whitehead Institute for Biomedical Research  
; APPLICANT: Lahn, Bruce  
; APPLICANT: Page, David  
; TITLE OF INVENTION: Genes in the No. 610386-Recombining Region of  
; FILE REFERENCE: WH197-08PA  
; CURRENT APPLICATION NUMBER: US/09/058,489  
; CURRENT FILING DATE: 1998-04-10  
; EARLIER APPLICATION NUMBER: 60/041,877  
; EARLIER FILING DATE: 1997-04-11  
; NUMBER OF SEQ ID NOS: 91  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 13  
; LENGTH: 5322  
; TYPE: DNA  
; ORGANISM: Human  
US-09-058-489-13

QY 1842 AACATGTTATCAATTTTGA 1860  
Db 2388 AACATGTTATCAATTTTGA 2406

RESULT 11  
US-09-214-808-1  
; Sequence 1, Application US/09214808A  
; Patent No. 6475793  
; GENERAL INFORMATION:  
; APPLICANT: Rosenthal, Andre  
; APPLICANT: Freiberg, Christoph  
; APPLICANT: Perret, Xavier Philippe  
; APPLICANT: Broughton, William John  
; TITLE OF INVENTION: Genomic Sequence of Rhizobium SP. NGR 234 Symbiotic  
; Patent No. 6475793  
; FILE REFERENCE: CARP0068  
; CURRENT APPLICATION NUMBER: US/09/214,808A  
; CURRENT FILING DATE: 1999-06-22  
; PRIOR APPLICATION NUMBER: PCT/IB97/00950  
; PRIOR FILING DATE: 1997-07-10  
; NUMBER OF SEQ ID NOS: 1  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 536165  
; TYPE: DNA  
; ORGANISM: Rhizobium  
US-09-214-808-1

Query Match 0.9%; Score 19; DB 4; Length 536165;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 GCGGCTATCGAGATGGA 422  
Db 146504 GCGGCTATCGAGATGGA 146522

RESULT 12  
US-08-651-155B-116  
; Sequence 116, Application US/08651155B  
; Patent No. 6365401  
; GENERAL INFORMATION:  
; APPLICANT: Mahan Dr., Michael J.  
; APPLICANT: Conner Mr., Christopher P.  
; APPLICANT: Hiethoff Mr., Douglas M.  
; TITLE OF INVENTION: METHOD AND PROBES FOR THE IDENTIFICATION  
; OF MICROBIAL GENES SPECIFICALLY INDUCED DURING HOST

TITLE OF INVENTION: INFECTION  
NUMBER OF SEQUENCES: 255  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Chairman, Byrum & Johnson, P.C.  
STREET: 1900 Fifteenth Street  
CITY: Boulder  
STATE: CO  
COUNTRY: USA  
ZIP: 80302  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/651,155B  
FILING DATE: 17-MAY-1996  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Petersen M., Steven C.  
REGISTRATION NUMBER: 36,238  
REFERENCE/DOCKET NUMBER: 17060.1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 303/546-1300  
TELEFAX: 303/449-5426  
TELEX: ABA1475  
INFORMATION FOR SEQ ID NO: 116:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 339 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: YES  
US-08-651-155B-116

Query Match 0.8%; Score 18; DB 4; Length 339;  
Best Local Similarity 100.0%; Pred. No. 51;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 CTCGAGCTTCATCATCAG 140  
Db 184 CTCGAGCTTCATCATCAG 201

RESULT 13  
US-09-194-036B-116  
Sequence 116, Application US/09194036B  
Patent No. 6548246  
GENERAL INFORMATION:  
APPLICANT: Mahan, Michael J.  
Conner, Christopher P.  
Hiechoff, Douglas W.  
TITLE OF INVENTION: METHOD AND PROBES FOR THE IDENTIFICATION  
OF MICROBIAL GENES SPECIFICALLY INDUCED DURING HOST  
INFECTION  
NUMBER OF SEQUENCES: 255  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Morrison & Foerster  
STREET: 755 Page Mill Road  
CITY: Mountain View  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/194,036B  
FILING DATE: 17-NO. 6548246-1998

CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US97/08208  
FILING DATE: 1997-05-16  
APPLICATION NUMBER: US 08/651,155  
FILING DATE: 1996-05-17  
ATTORNEY/AGENT INFORMATION:  
NAME: Shantanu Basu  
REGISTRATION NUMBER: 43,318  
REFERENCE/DOCKET NUMBER: 220002060601  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 813-5995  
TELEFAX: (650) 494-0792  
INFORMATION FOR SEQ ID NO: 116:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 339 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
HYPOTHETICAL: NO  
ANTI-SENSE: YES  
ORIGINAL SOURCE:  
ORGANISM: DNA (other)  
SEQUENCE DESCRIPTION: SEQ ID NO: 116:  
US-09-194-036B-116

Query Match 0.8%; Score 18; DB 4; Length 339;  
Best Local Similarity 100.0%; Pred. No. 51;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 CTCGAGCTTCATCATCAG 140  
Db 184 CTCGAGCTTCATCATCAG 201

RESULT 14  
US-09-222-575-116  
Sequence 116, Application US/09222575  
Patent No. 6387837  
GENERAL INFORMATION:  
APPLICANT: Yugu, Jjiang  
APPLICANT: Dillon, Davin C.  
APPLICANT: Mitcham, Jennifer L.  
APPLICANT: Xu, Jiangchun  
TITLE OF INVENTION: Compositions for the Treatment and Diagnosis of Breast Cancer  
FILE REFERENCE: 210121.470  
CURRENT APPLICATION NUMBER: US/09/222,575  
CURRENT FILING DATE: 1998-12-28  
NUMBER OF SEQ ID NOS: 174  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 116  
LENGTH: 421  
TYPE: DNA  
ORGANISM: Human  
US-09-222-575-116

Query Match 0.8%; Score 18; DB 4; Length 421;  
Best Local Similarity 100.0%; Pred. No. 51;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1382 GGTCGAGAAATGAAGAG 1399  
Db 12 GGTCGAGAAATGAAGAG 29

RESULT 15  
US-09-389-681-116  
Sequence 116, Application US/09389681A  
Patent No. 6518237  
GENERAL INFORMATION:  
APPLICANT: Yugu, Jjiang  
APPLICANT: Dillon, Davin C.

```
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.470C3
; CURRENT APPLICATION NUMBER: US/09/389,681A
; CURRENT FILING DATE: 1999-09-02
; NUMBER OF SEQ. ID NOS.: 463
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 116
; LENGTH: 421
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-389-681-116

Query Match          0.8%; Score 18; DB 4; Length 421;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1382 GGTCCAGAAATGAAGAAG 1399
    |||||
Db 12 GGTCCAGAAATGAAGAAG 29

Search completed: July 27, 2004, 18:11:16
Job time : 176 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 27, 2004, 16:33:16 ; Search time 999 Seconds  
(without alignments)  
10882.460 Million cell updates/sec

Title: US-09-714-865B-1

Perfect score: 2224  
Sequence: 1 acttgagtcacatcggggg.....aagtcgtgttgatgca 2224

Scoring table: OLIGO NUC  
Gapop 60.0 , Gapext 60.0

Searched: 3216467 seqs, 244149694 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6432934

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: listing first 45 summaries

Database :

Published Applications NA:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/PC7\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/PC7US\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*
- 16: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*
- 17: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*
- 18: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*
- 19: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1288	57.9	1984	16	US-10-104-047-1652 Sequence 1652, Ap
2	340	15.3	420	9	US-09-833-381-1475 Sequence 1475, Ap
3	278	12.5	591	10	US-09-764-891-1170 Sequence 1170, Ap
4	273	12.3	539	15	US-10-029-386-12207 Sequence 12207, A
5	271	12.2	271	15	US-10-029-386-25907 Sequence 25907, A
6	220	9.9	2194	10	US-09-764-891-7481 Sequence 7481, Ap
7	220	9.9	2195	10	US-09-764-891-7482 Sequence 7482, Ap
8	220	9.9	2197	10	US-09-764-891-7483 Sequence 7483, Ap
9	129	5.8	516	15	US-10-029-386-8746 Sequence 8746, Ap
10	128	5.8	153	15	US-10-029-386-22446 Sequence 22446, A
11	76	3.4	570	13	US-10-027-632-201125 Sequence 201125, A
12	76	3.4	570	16	US-10-027-632-201125 Sequence 201125, A
13	60	2.7	60	10	US-09-808-975-15365 Sequence 15365, A
14	36	1.6	1383	13	US-10-282-122A-41827 Sequence 41827, A

15	25	1.1	1335	16	US-10-369-493-24127	Sequence 24127, A
16	24	1.1	1347	16	US-10-369-493-24036	Sequence 24036, A
17	23	1.0	13020	16	US-10-408-168-1	Sequence 1, Appli
18	21	0.9	1842	9	US-09-815-242-6893	Sequence 6893, Ap
19	21	0.9	1842	13	US-10-282-122A-21926	Sequence 21926, A
20	21	0.9	1830121	15	US-10-329-960-1	Sequence 1, Appli
21	21	0.9	1830121	16	US-10-329-960-1	Sequence 1, Appli
22	20	0.9	620	13	US-10-027-632-8290	Sequence 8290, Ap
23	20	0.9	620	13	US-10-027-632-8291	Sequence 8291, Ap
24	20	0.9	620	16	US-10-027-632-8290	Sequence 8290, Ap
25	20	0.9	620	16	US-10-027-632-8291	Sequence 8291, Ap
26	20	0.9	633	13	US-10-027-632-22814	Sequence 22814, A
27	20	0.9	633	16	US-10-027-632-22816	Sequence 22816, A
28	20	0.9	637	13	US-10-027-632-214123	Sequence 214123, A
29	20	0.9	637	16	US-10-027-632-214123	Sequence 214123, A
30	20	0.9	704	13	US-10-027-632-98678	Sequence 98678, A
31	20	0.9	704	16	US-10-027-632-98678	Sequence 98678, A
32	20	0.9	710	13	US-10-027-632-163741	Sequence 163741, A
33	20	0.9	710	16	US-10-027-632-163741	Sequence 163741, A
34	20	0.9	1193	13	US-10-424-539-14492	Sequence 14492, A
35	20	0.9	2763	17	US-10-437-963-69857	Sequence 69857, A
36	20	0.9	2934	17	US-10-437-963-69859	Sequence 69859, A
37	20	0.9	9832	15	US-10-311-455-630	Sequence 630, Appli
38	20	0.9	96595	12	US-09-997-722-43	Sequence 43, Appli
39	20	0.9	122859	13	US-10-087-192-37	Sequence 37, Appli
40	19	0.9	133	15	US-10-029-386-23676	Sequence 23676, A
41	19	0.9	138	9	US-09-864-761-28835	Sequence 28835, A
42	19	0.9	367	13	US-10-424-539-16572	Sequence 16572, A
43	19	0.9	418	10	US-09-918-995-35717	Sequence 35717, A
44	19	0.9	510	13	US-10-027-632-69258	Sequence 69258, A
45	19	0.9	510	15	US-10-029-386-9976	Sequence 9976, Ap

## ALIGNMENTS

RESULT 1  
US-10-104-047-1652  
; Sequence 1652, Application US/10104047  
; Publication No. US20030236392A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: No. US20030236392A1 full length cDNA  
; FILE REFERENCE: H1-A0105  
; CURRENT APPLICATION NUMBER: US/10/104,047  
; PRIOR FILING DATE: 2002-03-25  
; PRIOR APPLICATION NUMBER:  
; NUMBER OF SEQ ID NOS: 4096  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1652  
; LENGTH: 1984  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-104-047-1652

Query Match	57.9%	Score 1288;	DB 16;	Length 1984;
Best Local Similarity	99.7%	Pred. No. 0;		
Matches 1488;	Conservative 0;	Mismatches 4;	Indels 0;	Gaps 0;
QY	733	AGAGCAAAAGTGACCTACATACCCCTCTCTCCACCTGAGATGAGAGCTCCATCTTGGC	792	
DB	292	AGAGCAAAAGTGACCTACATACCCCTCTCTCCACCTGAGATGAGAGCTCCATCTTGGC	351	
QY	793	ACATTATGACAGGAGATTAATCTTGACAAATATGACACTATCTTGGAGAGTCTGG	852	
DB	352	ACATTATGACAGGAGATTAATCTTGACAAATATGACACTATCTTGGAGAGTCTGG	411	
QY	853	ACATGATGACACGACGACATCTTGAGAGAGCTATCTGTGACACACTGAA	912	
DB	412	ACATGATGACACGACGACATCTTGAGAGAGCTATCTGTGACACACTGAA	471	
QY	913	TAAACAATTGTGTAAGCTGTTATACATGACTTCTCTGTGCAAAATACAGTATTC	972	

```

Db      472  TAAACATTTGCTAAAGCTGGTTATCTAGCTTACTCTGTCAGAAAAATATACAGTATTC 531
Qy      973  TATCATCTTGCAGAGAGATTTGATGGCTTGCTCAAAACAGGGCTCGGAGAGCTGC 1032
Db      532  TATCATCTTGCAGAGAGATTTGATGGCTTGCTCAAAACAGGGCTCGGAGAGCTGC 591
Qy      1033  GGGCTTTCTCTCAACCAATTTGGCTCATATGATGATGATGATGATGATGATGATGAT 1092
Db      592  GGGCTTTCTCTCAACCAATTTGGCTCATATGATGATGATGATGATGATGATGATGATGAT 651
Qy      1093  TAAAGATTTGAGAGAACAGAGTATTTATTTAGACCAACCTGAGAAATGGTCAACCA 1152
Db      652  TAAAGATTTGAGAGAACAGAGTATTTATTTAGACCAACCTGAGAAATGGTCAACCA 711
Qy      1153  GATTTATTTGAAAGCCAGAAAATTTCTTTGGAGCTTGTAAGAGCTGTTGTATATA 1212
Db      712  GATTTATTTGAAAGCCAGAAAATTTCTTTGGAGCTTGTAAGAGCTGTTGTATATA 771
Qy      1213  TGGGGGAAACCAAGCTGGAGATTCATTTGACAAATATGTAAGAGCTGTAATATATATG 1272
Db      772  TGGGGGAAACCAAGCTGGAGATTCATTTGACAAATATGTAAGAGCTGTAATATATATG 831
Qy      1273  TGCTACTCTGGAAGACTGATGATATCATAGGCAAGAAAGATTTGCTCAACAGAT 1332
Db      832  TGCTACTCTGGAAGACTGATGATATCATAGGCAAGAAAGATTTGCTCAACAGAT 891
Qy      1333  CAAATATCTTACTTTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1392
Db      892  CAAATATCTTACTTTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 951
Qy      1393  GAAGAGTTATATTTCTGGCCAGAGATGCTCAAGAAAGAGAGAGAGAGAGAGAGAGAGAG 1452
Db      952  GAAGAGTTATATTTCTGGCCAGAGATGCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1011
Qy      1453  CAGTGCAACTTTTCCAGAGAAATTCAGAGTTGGCTGCGAGAGTTTAAAGTCAATTA 1512
Db      1012  CAGTGCAACTTTTCCAGAGAAATTCAGAGTTGGCTGCGAGAGTTTAAAGTCAATTA 1071
Qy      1513  TCGCTTTGCTGCTGTTGGAAGAGGGTGGAGATGATGATGATGATGATGATGATGATGATGAT 1572
Db      1072  TCGCTTTGCTGCTGTTGGAAGAGGGTGGAGATGATGATGATGATGATGATGATGATGATGAT 1131
Qy      1573  CCAAGTTGGCCAGTTCTCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1632
Db      1132  CCAAGTTGGCCAGTTCTCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1191
Qy      1633  TGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1692
Db      1192  TGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1251
Qy      1693  TTGTCAAGAAAAAATATCACTACAGATATCAAGTATCGAGTGGGAAACAGAGAGAGAGAG 1752
Db      1252  TTGTCAAGAAAAAATATCACTACAGATATCAAGTATCGAGTGGGAAACAGAGAGAGAGAG 1311
Qy      1753  GCAAGCTCTTGGAGATTTTGGCTTGGAAAGTGGCCAGCTTCTGTTGCTTCACTCACTAGC 1812
Db      1312  GCAAGCTCTTGGAGATTTTGGCTTGGAAAGTGGCCAGCTTCTGTTGCTTCACTCACTAGC 1371
Qy      1813  TGGCAGAGGGGCTGATATTTGAAAAATGCAACAGTATCAATTTTGAATCTTCTCTAC 1872
Db      1372  TGGCAGAGGGGCTGATATTTGAAAAATGCAACAGTATCAATTTTGAATCTTCTCTAC 1431
Qy      1873  CATTGATGAATATGTTGATGAAATGGGCTGATGATGATGATGATGATGATGATGATGATGAT 1932
Db      1432  CATTGATGAATATGTTGATGAAATGGGCTGATGATGATGATGATGATGATGATGATGATGAT 1491
Qy      1933  AATTTCCTTTTGAATCTTGAATCGGTAACCTTTAGCAACAGCTCTAGTAAAGATAT 1992
Db      1492  AATTTCCTTTTGAATCTTGAATCGGTAACCTTTAGCAACAGCTCTAGTAAAGATAT 1551
Qy      1993  GACAGATGCTCAACAGAGATGTTCTGCAATGTTGGAAGAAATTCCTTTAGTACATACAT 2052

```

```

Db      1552  GACAGATGCTCAACAGAGATGTTCTGCAATGTTGGAAGAAATTTGCTTTAGTACATACAT 1611
Qy      2053  TCCTGGCTTCAGTGGTATGATCAAGAGAGAAAGTGTGTCATGAGTTGATATCCAGAAAGG 2112
Db      1612  TCCTGGCTTCAGTGGTATGATCAAGAGAGAAAGTGTGTCATGAGTTGATATCCAGAAAGG 1671
Qy      2113  CAAAGAGACTTTGAACACAGTGGGTTTCTGTTCCGACGAGCTCCCAATCAGATATGA 2172
Db      1672  CAAAGAGACTTTGAACACAGTGGGTTTCTGTTCCGACGAGCTCCCAATCAGATATGA 1731
Qy      2173  TGAATCATGGGATTTAAAGCCAAACATCTTCAAGTGTGTTGATATCA 2224
Db      1732  TGAATCATGGGATTTAAAGCCAAACATCTTCAAGTGTGTTGATATCA 1783

```

## RESULT 2

```

US-09-833-381-1475
; Sequence 1475, Application US/09833381
; Patent No. US20020132090A1
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: No. US20020132090A1el Nucleic Acid and Protein Homologs
; FILE REFERENCE: 5800-119
; CURRENT APPLICATION NUMBER: US/09/833,381
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 09/516,448
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 2050
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1475
; LENGTH: 420
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-833-381-1475

```

## Query Match 15.3%; Score 340; DB 9; Length 420;

```

Best Local Similarity 99.7%; Pred. No. 4.9e-11;
Matches 390; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      820  CAAATACAGACTATTTCTTGTGAGAGTGTGACATGATGACACACAGCAATTCAGAC 879
Db      30  CAAATACAGACTATTTCTTGTGAGAGTGTGACATGATGACACACAGCAATTCAGAC 89
Qy      880  TTTTGAAGAGCTATCTCTGTGACAGACTGAAATTAACAACATTTGCTTAAGCTGTTATAC 939
Db      90  TTTTGAAGAGCTATCTCTGTGACAGACTGAAATTAACAACATTTGCTTAAGCTGTTATAC 149
Qy      940  TAAGCTTACTCTGTGCAAAAAATACAGATTCCTATCATACTTGCAGAGAGAGATTGAT 999
Db      150  TAAGCTTACTCTGTGCAAAAAATACAGATTCCTATCATACTTGCAGAGAGAGATTGAT 209
Qy      1000  GGGTTGTGCTCAACAGGGTCTGGAGAAAGACTGGGGCTTTTCTCTACCAATTTTGGCTCA 1059
Db      210  GGGTTGTGCTCAACAGGGTCTGGAGAAAGACTGGGGCTTTTCTCTACCAATTTTGGCTCA 269
Qy      1060  TATGATGATGATGGAATTAATGCGCAGTCTTTAAAGAGTTGAGAGACAGAGTGTAT 1119
Db      270  TATGATGATGATGGAATTAATGCGCAGTCTTTAAAGAGTTGAGAGACAGAGTGTAT 329
Qy      1120  TATTGTACACCAACTCGAGATTTGTCACACAGATTTATTGGAAGCCAGAAATTTTC 1179
Db      330  TATTGTACACCAACTCGAGATTTGTCACACAGATTTATTGGAAGCCAGAAATTTTC 389
Qy      1180  TTTTGGAGCTTGTGTAAGAGCTGTTGTTATA 1210
Db      390  TTTTGGAGCTTGTGTAAGAGCTGTTGTTATA 420

```

## RESULT 3

```

US-09-764-891-1170
; Sequence 1170, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:

```

APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PC006  
CURRENT APPLICATION NUMBER: US/09/764,891  
CURRENT FILING DATE: 2001-01-17  
Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 10231  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1170  
LENGTH: 591  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-764-891-1170

Query Match 12.5%; Score 278; DB 10; Length 591;

Best Local Similarity 99.7%; Pred. No. 9,38-138; Mismatches 1; Indels 0; Gaps 0;

Matches 328; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1896 TTGGGCGTACTGTCGTTGGGAATCTGCGACAGCAATTTCTTTTGTATCTTGAT 1955

Db 30 TTGGGCGTACTGTCGTTGGGAATCTGCGACAGCAATTTCTTTTGTATCTTGAT 89

QY 1956 CGGATTAACATTTAGCACACGCTCTAGTAAAGTATTGACAGATGCTCAACAGATGTC 2015

Db 90 CGGATTAACATTTAGCACACGCTCTAGTAAAGTATTGACAGATGCTCAACAGATGTC 149

QY 2016 CTGCATGTTGAGAGAAATTCCTTTAGTACATATCTTCTGCGCTTCACTGATGATCA 2075

Db 150 CTGCATGTTGAGAGAAATTCCTTTAGTACATATCTTCTGCGCTTCACTGATGATCA 209

QY 2076 GAGGAAAAGTGTTCATCATGTTGATACCAAGAAAGGCAAGACATTTGAAACAGCTG 2135

Db 210 GAGGAAAAGTGTTCATCATGTTGATACCAAGAAAGGCAAGACATTTGAAACAGCTG 269

QY 2136 GGTTCCTTCTTCACGAGCTCCCAATCCAGTAGTATGATGATGATGATGATGATGAT 2195

Db 270 GGTTCCTTCTTCACGAGCTCCCAATCCAGTAGTATGATGATGATGATGATGATGAT 329

QY 2196 ACATCCTTCAAGTCTGTGTTTGTATGCA 2224

Db 330 ACATCCTTCAAGTCTGTGTTTGTATGCA 358

RESULT 4  
US-10-029-386-12207/c  
Sequence 12207, Application US/10029386  
Publication No. US20030194704A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Hanzel, David K.  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: AEWOMICA-X-2  
CURRENT APPLICATION NUMBER: US/10/029,386  
CURRENT FILING DATE: 2001-12-20  
NUMBER OF SEQ ID NOS: 34288  
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 12207  
LENGTH: 539  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC016639.5  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.98  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.52  
OTHER INFORMATION: EST HUMAN HIT: AA399611.1, EVALUE 1.00e-64  
OTHER INFORMATION: NT HIT: g116156979, EVALUE 0.00e+00  
OTHER INFORMATION: SWISSPROT HIT: Q61496, EVALUE 5.00e-45  
US-10-029-386-12207

Query Match 12.3%; Score 273; DB 15; Length 539;  
Best Local Similarity 100.0%; Pred. No. 4,4e-135;

Matches 273; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1728 GTGATCGGGAACAGAGAGCGGAGCAAGCTCTTGAGATTTTGGCTTTGAAAGTGC 1787

Db 507 GTGATCGGGAACAGAGAGCGGAGCAAGCTCTTGAGATTTTGGCTTTGAAAGTGC 448

QY 1788 CAGTCTTGTGCTACTAGTACGCTGCGAGAGGCTGATTTGAAATGCAACATG 1847

Db 447 CAGTCTTGTGCTACTAGTACGCTGCGAGAGGCTGATTTGAAATGCAACATG 388

QY 1848 TTATCAATTTGATCTTCTTACCATGATGATATGTTGATGATGATGATGATGATGAT 1907

Db 387 TTATCAATTTGATCTTCTTACCATGATGATATGTTGATGATGATGATGATGATGAT 328

QY 1908 GTGCTGTGGGAATCTGCGACAGCAATTTCTTTTGTATCTTGATGATGATGATGAT 1967

Db 327 GTGCTGTGGGAATCTGCGACAGCAATTTCTTTTGTATCTTGATGATGATGATGAT 268

QY 1968 TAGCAGGCTCTAGTAAAGTATTGACAGATG 2000

Db 267 TAGCAGGCTCTAGTAAAGTATTGACAGATG 235

RESULT 5  
US-10-029-386-25907/c  
Sequence 25907, Application US/10029386  
Publication No. US20030194704A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Hanzel, David K.  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: AEWOMICA-X-2  
CURRENT APPLICATION NUMBER: US/10/029,386  
CURRENT FILING DATE: 2001-12-20  
NUMBER OF SEQ ID NOS: 34288  
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 25907  
LENGTH: 271  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC016639.5  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.98  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.52  
OTHER INFORMATION: EST HUMAN HIT: AA399611.1, EVALUE 2.00e-64  
OTHER INFORMATION: NT HIT: g116156979, EVALUE 0.00e+00  
OTHER INFORMATION: SWISSPROT HIT: Q61496, EVALUE 2.00e-45  
US-10-029-386-25907

Query Match 12.2%; Score 271; DB 15; Length 271;  
Best Local Similarity 100.0%; Pred. No. 4,9e-134;  
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1729 TGATCGGGAACAGAGAGCGGAGCAAGCTCTTGAGATTTTGGCTTTGAAAGTGC 1788

Db 271 TGATCGGGAACAGAGAGCGGAGCAAGCTCTTGAGATTTTGGCTTTGAAAGTGC 212

QY 1789 AGTTCCTTGTGCTACTAGTACGCTGCGAGAGGCTGATTTGAAATGCAACATG 1848

Db 211 AGTTCCTTGTGCTACTAGTACGCTGCGAGAGGCTGATTTGAAATGCAACATG 152

QY 1849 TATCAATTTGATCTTCTTACCATGATGATATGTTGATGATGATGATGATGATGAT 1908

Db 151 TATCAATTTGATCTTCTTACCATGATGATATGTTGATGATGATGATGATGATGAT 92

QY 1909 TCGTGTGGGAATCTGCGACAGCAATTTCTTTTGTATCTTGATGATGATGATGAT 1968

Db 91 TCGTGTGGGAATCTGCGACAGCAATTTCTTTTGTATCTTGATGATGATGATGATGAT 32

QY 1969 AGCAGGCTCTAGTAAAGTATTGACAGAT 1999

|||



Db 31 AGCAGACGCTCTAGTAAAGTATTGACAGAT 1

RESULT 6

US-09-764-891-7481  
Sequence 7481, Application US/09764891  
Publication No. US20030077808A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

FILE REFERENCE: PC006

CURRENT APPLICATION NUMBER: US/09/764,891

CURRENT FILING DATE: 2001-01-17

Prior application data removed - consult PALM or file wrapper

NUMBER OF SEQ ID NOS: 10231

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 7481

LENGTH: 2194

TYPE: DNA

ORGANISM: Homo sapiens

US-09-764-891-7481

Query Match

Best Local Similarity 100.0%; Pred. No. 1.4e-106; Length 2194;  
Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1781 AAGTCCCGAGTCTCTGCTGCTACTTCACTGAGTCCAGAGGCTGGATATTGAAAATGTG 1840

1 AAGTCCCGAGTCTCTGCTGCTACTTCACTGAGTCCAGAGGCTGGATATTGAAAATGTG 60

Db 1841 CAACATGTTATCAATTTGATCTTCTTACCATGATGATATGTTATCGAATTGGG 1900

61 CAACATGTTATCAATTTGATCTTCTTACCATGATGATATGTTATCGAATTGGG 120

Db 1901 CGTACTGCTGCTGTGGGAAATCTGCGAGAGCAATTTCTTTTGAATCTTGATCGAT 1960

121 CGTACTGCTGCTGTGGGAAATCTGCGAGAGCAATTTCTTTTGAATCTTGATCGAT 180

Db 1961 AACCATTTAGACAGCCTCTAGTAAAGTATTGACAGATG 2000

181 AACCATTTAGACAGCCTCTAGTAAAGTATTGACAGATG 220

RESULT 7  
US-09-764-891-7482  
Sequence 7482, Application US/09764891  
Publication No. US20030077808A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PC006  
CURRENT APPLICATION NUMBER: US/09/764,891  
CURRENT FILING DATE: 2001-01-17  
Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 10231  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 7482  
LENGTH: 2195  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-764-891-7482

Query Match

Best Local Similarity 100.0%; Pred. No. 1.4e-106; Length 2195;  
Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1781 AAGTCCCGAGTCTCTGCTGCTACTTCACTGAGTCCAGAGGCTGGATATTGAAAATGTG 1840

1 AAGTCCCGAGTCTCTGCTGCTACTTCACTGAGTCCAGAGGCTGGATATTGAAAATGTG 60

Db 1841 CAACATGTTATCAATTTGATCTTCTTACCATGATGATATGTTATCGAATTGGG 1900

61 CAACATGTTATCAATTTGATCTTCTTACCATGATGATATGTTATCGAATTGGG 120

Db 1901 CGTACTGCTGCTGTGGGAAATCTGCGAGAGCAATTTCTTTTGAATCTTGATCGAT 1960

121 CGTACTGCTGCTGTGGGAAATCTGCGAGAGCAATTTCTTTTGAATCTTGATCGAT 180

Db 1961 AACCATTTAGACAGCCTCTAGTAAAGTATTGACAGATG 2000

181 AACCATTTAGACAGCCTCTAGTAAAGTATTGACAGATG 220

RESULT 8  
US-09-764-891-7483  
Sequence 7483, Application US/09764891  
Publication No. US20030077808A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PC006  
CURRENT APPLICATION NUMBER: US/09/764,891  
CURRENT FILING DATE: 2001-01-17  
Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 10231  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 7483  
LENGTH: 2197  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-764-891-7483

Query Match  
Best Local Similarity 100.0%; Pred. No. 1.4e-106; Length 2197;  
Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1781 AAGTCCCGAGTCTCTGCTGCTACTTCACTGAGTCCAGAGGCTGGATATTGAAAATGTG 1840

1 AAGTCCCGAGTCTCTGCTGCTACTTCACTGAGTCCAGAGGCTGGATATTGAAAATGTG 60

Db 1841 CAACATGTTATCAATTTGATCTTCTTACCATGATGATATGTTATCGAATTGGG 1900

61 CAACATGTTATCAATTTGATCTTCTTACCATGATGATATGTTATCGAATTGGG 120

Db 1901 CGTACTGCTGCTGTGGGAAATCTGCGAGAGCAATTTCTTTTGAATCTTGATCGAT 1960

121 CGTACTGCTGCTGTGGGAAATCTGCGAGAGCAATTTCTTTTGAATCTTGATCGAT 180

Db 1961 AACCATTTAGACAGCCTCTAGTAAAGTATTGACAGATG 2000

181 AACCATTTAGACAGCCTCTAGTAAAGTATTGACAGATG 220

RESULT 9  
US-10-029-386-8746/C  
Sequence 8746, Application US/10029386  
Publication No. US20030194704A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Hanzel, David K.  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: AEMICA-X-2  
CURRENT APPLICATION NUMBER: US/10/029,386  
CURRENT FILING DATE: 2001-12-20  
NUMBER OF SEQ ID NOS: 34288  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 8746  
LENGTH: 516  
TYPE: DNA  
ORGANISM: Homo sapiens  
OTHER INFORMATION: MAP TO AC016639.5  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.4

Qy 1901 CGTACTGCTGCTGTGGGAAATCTGCGAGAGCAATTTCTTTTGAATCTTGATCGAT 1960  
Db 121 CGTACTGCTGCTGTGGGAAATCTGCGAGAGCAATTTCTTTTGAATCTTGATCGAT 180  
Qy 1961 AACCATTTAGACAGCCTCTAGTAAAGTATTGACAGATG 2000  
Db 181 AACCATTTAGACAGCCTCTAGTAAAGTATTGACAGATG 220

RESULT 9

US-10-029-386-8746/C  
Sequence 8746, Application US/10029386  
Publication No. US20030194704A1

GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.

APPLICANT: Hanzel, David K.

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

FILE REFERENCE: AEMICA-X-2

CURRENT APPLICATION NUMBER: US/10/029,386

CURRENT FILING DATE: 2001-12-20

NUMBER OF SEQ ID NOS: 34288

SOFTWARE: Anomax Sequence Listing Engine vers. 1.1

SEQ ID NO 8746

LENGTH: 516

TYPE: DNA

ORGANISM: Homo sapiens

US-10-029-386-8746/C

Query Match

Best Local Similarity 100.0%; Pred. No. 1.4e-106; Length 2197;  
Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1781 AAGTCCCGAGTCTCTGCTGCTACTTCACTGAGTCCAGAGGCTGGATATTGAAAATGTG 1840

1 AAGTCCCGAGTCTCTGCTGCTACTTCACTGAGTCCAGAGGCTGGATATTGAAAATGTG 60

Db 1841 CAACATGTTATCAATTTGATCTTCTTACCATGATGATATGTTATCGAATTGGG 1900

61 CAACATGTTATCAATTTGATCTTCTTACCATGATGATATGTTATCGAATTGGG 120

Db 1901 CGTACTGCTGCTGTGGGAAATCTGCGAGAGCAATTTCTTTTGAATCTTGATCGAT 1960

121 CGTACTGCTGCTGTGGGAAATCTGCGAGAGCAATTTCTTTTGAATCTTGATCGAT 180

Db 1961 AACCATTTAGACAGCCTCTAGTAAAGTATTGACAGATG 2000

181 AACCATTTAGACAGCCTCTAGTAAAGTATTGACAGATG 220

RESULT 10  
US-10-029-386-8746/C  
Sequence 8746, Application US/10029386  
Publication No. US20030194704A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Hanzel, David K.  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: AEMICA-X-2  
CURRENT APPLICATION NUMBER: US/10/029,386  
CURRENT FILING DATE: 2001-12-20  
NUMBER OF SEQ ID NOS: 34288  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 8746  
LENGTH: 516  
TYPE: DNA  
ORGANISM: Homo sapiens  
OTHER INFORMATION: MAP TO AC016639.5  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.4



Qy	1358	GATCGCATGTTGGATATGGGTTTG	1382
Db	478	GATCCGATGTTGGATATGGGTTTG	502

Wed Jul 28 11:42:29 2004

Search completed: July 27, 2004, 20:40:25  
Job time : 1006 secs

us-09-714-865b-1.rnpb

***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 27, 2004, 12:56:43 ; Search time 5696 Seconds  
(without alignments)  
11659.672 Million cell updates/sec

Title: US-09-714-865B-1

Perfect score: 2224  
Sequence: 1 actgaagtcaccatggggg.....aagtcgtggtttgatgca 2224

Scoring table: OLIGO NUC  
Gapop 60.0 , Gapexc 60.0

Searched: 27513269 seqs, 14931090276 residues

Word size : 0

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_estbm:\*  
3: em\_estin:\*  
4: em\_esttm:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hc1:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hc1:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vit:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gset:\*  
29: gb\_gset2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1829	82.2	2523	11	BC030638 Homo sapi
2	665	29.9	850	14	CD251828 AGENCOURT
3	626	28.1	896	14	CD57531 AGENCOURT
4	616	27.7	682	12	BC718274 AGENCOURT

Result No.	Score	Query Match	Length	ID	Description
5	616	27.7	693	12	B1560849
6	612	27.5	798	12	BG717110
7	568	25.5	885	14	CD557160
8	544	24.5	981	14	CD359812
9	539	24.1	868	14	CD358073
10	535	24.1	792	13	BQ429126
11	521	23.4	914	14	CD557137
12	509	22.9	747	12	BG719449
13	480	21.6	657	12	BG717812
14	471	21.2	933	14	CD557866
15	469	21.1	651	12	BG717354
16	441	19.8	822	12	B1464267
17	440	19.8	846	12	BG718634
18	430	19.3	861	12	B1461393
19	422	19.0	830	12	B1462207
20	413	18.6	699	12	BG718822
21	412	18.5	743	12	B1463502
22	382	17.6	841	12	B1458115
23	388	17.4	939	14	CD300339
24	381	17.1	884	12	B1560499
25	361	16.2	918	12	B1561001
26	337	15.2	810	12	B1463846
27	326	14.7	758	12	B1559479
28	321	14.4	482	9	AL042306
29	300	13.5	497	13	BK282689
30	300	13.5	555	9	AA399611
31	269	12.1	555	9	AI217144
32	267	12.0	847	12	B1562624
33	246	11.1	917	12	B1463447
34	245	11.0	805	14	CB230120
35	232	10.9	491	9	AA398976
36	223	10.5	335	9	AA383535
37	225	10.1	347	9	AI953070
38	204	9.2	700	12	BG718861
39	161	7.2	604	12	B1460628
40	155	7.0	300	9	AI337133
41	155	7.0	300	9	AI654417
42	155	7.0	300	10	BBS50307
43	155	7.0	300	10	BBS50448
44	152	6.8	690	14	CD767258
45	137	6.2	360	10	AW449170

## ALIGNMENTS

RESULT 1  
BC030638  
LOCUS  
DEFINITION  
Homo sapiens CDNA clone IMAGE:4822525, containing frame-shift errors.  
ACCESSION  
BC030638  
VERSION  
BC030638.1 GI:21040391  
KEYWORDS  
HTC.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE  
1 (bases 1 to 2523)  
Straussberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenner,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.S., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Uebli,T.B., Toshylyki,S., Carninci,P., Prange,C., Rana,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahy,U., Hellon,E., Ketterman,M., Madan,A., Rodriguez,S.,

TITLE  
JOURNAL  
PUBLISHED  
REFERENCE  
AUTHORS  
JOURNAL  
TITLE  
SUBMITTED (20-MAY-2002) National Institutes of Health, Mammalian  
Gene Collection (MGC). Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki  
Toshiyuki and Piero Carninci (RIKEN)  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)  
DNA Sequencing by: Institute for Systems Biology  
<http://www.systemsbio.org>  
Contact: [amadamsystemsbiology.org](mailto:amadamsystemsbiology.org)  
Anup Madan, Jessica Fahy, Erin Helton, Mark Kettelman, Anuradha  
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

REMARK  
COMMENT

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LINL at: <http://image.lnl.gov>  
Series: IRAX Plate: 33 Row: d Column: 8  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 9507226  
This clone has the following problem: frame shifted.

FEATURES  
SOURCE

1. 2523  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4822528"  
/tissue\_type="Testis"  
/clone\_lib="NIH MGC\_97"  
/lab\_host="DH10B"  
/note="Vector: pBluescript"

ORIGIN

Query Match 82.2%; Score 1829; DB 11; Length 2523;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 2199; Conservative 0; Mismatches 5; Indels 1; Gaps 1;  
QY 20 GATGAGATTGGAGAGAGAAATCAACCTCATATGCTTCTCATGTTCCCATATTTGAG 79  
DB 104 GATGAGATTGGAGAGAGAAATCAACCTCATATGCTTCTCATGTTCCCATATTTGAG 163  
QY 80 AAGGATGCTATTTCTGGAGAAATGAGAGACAAATTTTAAACAGACCTCAGCTCATATCA 139  
DB 164 AAGGATGCTATTTCTGGAGAAATGAGAGACAAATTTTAAACAGACCTCAGCTCATATCA 223  
QY 140 GAATGAGATGAGACCTTCTCGAAGAGATCATTTGATGAAAGATTTGCTCTGGG 199  
DB 224 GAATGAGATGAGACCTTCTCGAAGAGATCATTTGATGAAAGATTTGCTCTGGG 283  
QY 200 CGGAATTTTGAACACAGATGCTGATGATTAATAGGAGATTAATACATCCAAAG 259  
DB 284 CGGAATTTTGAACACAGATGCTGATGATTAATAGGAGATTAATACATCCAAAG 343  
QY 260 GGTGCTTTTGAAGTGAAGAGTTTGAACACAGAGTTTTCACACAGAGTTTGA 319  
DB 344 GGTGCTTTTGAAGTGAAGAGTTTGAACACAGAGTTTTCACACAGAGTTTGA 403  
QY 320 GATGCTATAGCTCTGCTTTCTGAGAGAGCTGATGATGATGCTGCAAGATTAATCA 379  
DB 404 GATGCTATAGCTCTGCTTTCTGAGAGAGCTGATGATGATGCTGCAAGATTAATCA 463

QY 380 CGAAACAGAGGCTTTTCCAAAGAGAGGCGCTATCGAATGGAATTAATTCAGAGCTTCA 439  
DB 464 CGAAACAGAGGCTTTTCCAAAGAGAGGCGCTATCGAATGGAATTAATTCAGAGCTTCA 523  
QY 440 GGGCCATACAGAGAGGAGAGAGAGTATTTCCGAGGTTCCGAGAGATTTGCTCTA 499  
DB 524 GGGCCATACAGAGAGGAGAGAGAGTATTTCCGAGGTTCCGAGAGATTTGCTCTA 583  
QY 500 GGAAGTCCAAATTAATGACTTAGACCCAGACGATGATGACAGGCACTGCTGCTTTT 559  
DB 584 GGAAGTCCAAATTAATGACTTAGACCCAGACGATGATGACAGGCACTGCTGCTTTT 643  
QY 560 GGTCTGAAAGACAGATTAATGAGGACAGGATGATGATGATGATGATGATGATGATGAT 619  
DB 644 GGTCTGAAAGACAGATTAATGAGGACAGGATGATGATGATGATGATGATGATGATGAT 703  
QY 620 GCGAGTGAAGTGAAGAGAGGAGTGTATTAACAAAGTTTAATGAAGATTAATGAGGCTCT 679  
DB 704 GCGAGTGAAGTGAAGAGAGGAGTGTATTAACAAAGTTTAATGAAGATTAATGAGGCTCT 763  
QY 680 GGAAGATTTTGGAGAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 739  
DB 764 GGAAGATTTTGGAGAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 823  
QY 740 AAGTGAAGTGAAG 799  
DB 824 AAGTGAAGTGAAG 883  
QY 800 CAGACAGGATTAATCTTGCACAAATAGACACATTTCTTGGAGAGTGTGACAGAT 859  
DB 884 CAGACAGGATTAATCTTGCACAAATAGACACATTTCTTGGAGAGTGTGACAGAT 943  
QY 860 GCACACAGGATTAATCTTGCACAAATAGACACATTTCTTGGAGAGTGTGACAGAT 919  
DB 944 GCACACAGGATTAATCTTGCACAAATAGACACATTTCTTGGAGAGTGTGACAGAT 1003  
QY 920 ATTGCTAAAGCTGTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 979  
DB 1004 ATTGCTAAAGCTGTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1063  
QY 980 CTTGAG 1039  
DB 1064 CTTGAG 1123  
QY 1040 CTCTACCAATTTTGGCTCATATGATGATGATGATGATGATGATGATGATGATGATGAT 1099  
DB 1124 CTCTACCAATTTTGGCTCATATGATGATGATGATGATGATGATGATGATGATGATGAT 1183  
QY 1100 TTGACAG 1159  
DB 1184 TTGACAG 1243  
QY 1160 TTGAG 1219  
DB 1244 TTGAG 1303  
QY 1220 ACCGAGCTGGACATTTCAATTTGACAAATAGTACAGAGCTGTATTAATTAATGCTACT 1279  
DB 1304 ACCGAGCTGGACATTTCAATTTGACAAATAGTACAGAGCTGTATTAATTAATGCTACT 1363  
QY 1280 COTGAG 1339  
DB 1364 CC-GGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1422  
QY 1340 TTAGTTTGAAGTGAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1399  
DB 1423 TTAGTTTGAAGTGAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1482  
QY 1400 TTAATTTTCCAG 1459  
DB 1483 TTAATTTTCCAG 1542



QY 1460 ACTTTTCCAGAGGAATAATTCAGAGGTGGCTGCAGAGTCTTTTAAAGTCAAATTAATCTGTTT 1519  
Db 1543 ACTTTTCCAGAGGAATAATTCAGAGGTGGCTGCAGAGTCTTTTAAAGTCAAATTAATCTGTTT 1602  
QY 1520 GTTGCCTTTGACAGAGGTGGCTGCAGAGTCTTTTAAAGTCAAATTAATCTGTTT 1579  
Db 1603 GTTGCCTTTGACAGAGGTGGCTGCAGAGTCTTTTAAAGTCAAATTAATCTGTTT 1662  
QY 1580 GGCACAGTCTCAAAAAGAGAAAAGCTGCTGAAATTCGCGAACAATAGGGAGTGAAGA 1639  
Db 1663 GGCACAGTCTCAAAAAGAGAAAAGCTGCTGAAATTCGCGAACAATAGGGAGTGAAGA 1722  
QY 1640 ACTATGCTCTTTGTTGTAAGTCTGTAAGAAAAGAGATTTTACTGCACTTTTCTTGTCA 1699  
Db 1723 ACTATGCTCTTTGTTGTAAGTCTGTAAGAAAAGAGATTTTACTGCACTTTTCTTGTCA 1782  
QY 1700 GAAAAAATATCACTACAGAGTATCCATGATCGGAGACAGAGAGCGGAGCAAGCT 1759  
Db 1783 GAAAAAATATCACTACAGAGTATCCATGATCGGAGACAGAGAGCGGAGCAAGCT 1842  
QY 1760 CTTCGAGATTTTTCGCTTTGAGAAAAGTGCAGATTTCTGTTGCTACTTCAAGTCTGCCA 1819  
Db 1843 CTTCGAGATTTTTCGCTTTGAGAAAAGTGCAGATTTCTGTTGCTACTTCAAGTCTGCCA 1902  
QY 1820 GGGCTGATATTTGAAAATGTCAGAAAGTATGATTTGATCTTCTCTTACATTTGAT 1879  
Db 1903 GGGCTGATATTTGAAAATGTCAGAAAGTATGATTTGATCTTCTCTTACATTTGAT 1962  
QY 1880 GAATATGTTGATGAAATTTGGCGGATGCTGCTGTTGAGAAATCTGGAGAGCAATTTCC 1939  
Db 1963 GAATATGTTGATGAAATTTGGCGGATGCTGCTGTTGAGAAATCTGGAGAGCAATTTCC 2022  
QY 1940 TTTTGTGATCTTGAATGGAATTAACCATTTAGCACAGCTTCTAGTAAAGTATTAAGAT 1999  
Db 2023 TTTTGTGATCTTGAATGGAATTAACCATTTAGCACAGCTTCTAGTAAAGTATTAAGAT 2082  
QY 2000 GCTCAACAGAGATGTTCTGATGAGTGGTGAAGAAATTCCTTTAGTACATCTTCTGCG 2059  
Db 2083 GCTCAACAGAGATGTTCTGATGAGTGGTGAAGAAATTCCTTTAGTACATCTTCTGCG 2142  
QY 2060 TTCAATGATGATGACAGAGAAAGTGTTCATCACTGATGATACCAAGAAAGGCAAGAG 2119  
Db 2143 TTCAATGATGATGACAGAGAAAGTGTTCATCACTGATGATACCAAGAAAGGCAAGAG 2202  
QY 2120 ACTTTGAACAGAGTGGTGTTCCTTCTCAAGAGCTCCCAATCCAGAGATGATGATGATCA 2179  
Db 2203 ACTTTGAACAGAGTGGTGTTCCTTCTCAAGAGCTCCCAATCCAGAGATGATGATGATCA 2262  
QY 2180 TGGGATTTAAAGCCAAAACATCTCTCAAGTCTGTTGTTGATGCA 2224  
Db 2263 TGGGATTTAAAGCCAAAACATCTCTCAAGTCTGTTGTTGATGCA 2307

RESULT 2  
CD251828 850 bp mRNA linear EST 22-MAY-2003  
LOCUS AGENCOURT 14206017 NIH MGC 180 Homo sapiens cDNA clone  
DEFINITION IMAGE:30383198 5', mRNA sequence.  
ACCESSION CD251828  
VERSION CD251828.1 GI:31012294  
SOURCE EST.  
ORGANISM Homo sapiens (human)  
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE NIH-MGC <http://mgs.nci.nih.gov/>.  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE Unpublished (1999)  
JOURNAL Contact: Daniela S. Gerhard, Ph.D.  
COMMENT Office of Cancer Genomics  
National Cancer Institute / NIH  
Ridg 31 Em0407 Bethesda MD 20892

FEATURES  
source  
Email: [cgapbs-td@mail.nih.gov](mailto:cgapbs-td@mail.nih.gov)  
Tissue Procurement: Dr. Michael Brownstein  
CDNA Library Preparation: Invitrogen Corp  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMU)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LMU at:  
<http://image.llnl.gov>  
Plate: ND4M446 row: k column: 15  
High quality sequence stop: 626.  
Location/Qualifiers  
1..850  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:30383198"  
/lab\_host="DH10B-Ton A (T1 and T5 phage resistances)"  
/note="Organ: Testis; Vector: pCMV-Sport6.1; Site 1: NotI;  
Site 2: EcoRV (destroyed); Library is oligo-dT primed and  
directionally cloned (EcoRV site is destroyed upon  
cloning). Average insert size 1.68 kb. Library was  
constructed by (Invitrogen). Note: this is a NIH\_MGC  
Library."

ORIGIN  
Query Match 29.9%; Score 665; DB 14; Length 850;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 665; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 20 GATGAAGATTGGGAAGCAGAAATCAACCTCATATGCTTCCTCCATATTTAG 79  
Db 17 GATGAAGATTGGGAAGCAGAAATCAACCTCATATGCTTCCTCCATATTTAG 76  
QY 80 AAGATAGATATTCGAGAAAATGAGACAATTTTAAACAGAGCTCCAGCTTCATCATCA 139  
Db 77 AAGATAGATATTCGAGAAAATGAGACAATTTTAAACAGAGCTCCAGCTTCATCATCA 136  
QY 140 GAATGATGATGAGACCTCTTCGAAAGATATTTCAAGAAAAGTGGATTCCTCTGAG 199  
Db 137 GAATGATGATGAGACCTCTTCGAAAGATATTTCAAGAAAAGTGGATTCCTCTGAG 196  
QY 200 CGGAATTTTGGAAACAGAGATGCTGATGATGATTAAGAGAGATATATCCAAATG 259  
Db 197 CGGAATTTTGGAAACAGAGATGCTGATGATGATTAAGAGAGATATATCCAAATG 256  
QY 260 GGTGTTTGGAGTTGGAAAAGATTTTGGAAAACAGAGTTTTCAAAACAGAGTTTGA 319  
Db 257 GGTGTTTGGAGTTGGAAAAGATTTTGGAAAACAGAGTTTTCAAAACAGAGTTTGA 316  
QY 320 GATGGATGATGCTGTTTCTGAGAGAGTCTAGTATATGACTGCGAAGATATCCACA 379  
Db 317 GATGGATGATGCTGTTTCTGAGAGAGTCTAGTATATGACTGCGAAGATATCCACA 376  
QY 380 CGGAACAGAGGTTTCCAGAGAGCGGCTATGAGATGGAATATATTCAGAACTTCA 439  
Db 377 CGGAACAGAGGTTTCCAGAGAGCGGCTATGAGATGGAATATATTCAGAACTTCA 436  
QY 440 GGGCCATTCAGAGAGGTGGAAGAGTGTGTTCCAGAGTGGCGGAGAGTTTGGTGA 499  
Db 437 GGGCCATTCAGAGAGGTGGAAGAGTGTGTTCCAGAGTGGCGGAGAGTTTGGTGA 496  
QY 500 GGAAGTCCAAATATGACTTATGACCCAGACGAATGATGACGCGCATGTGGCCTTTT 559  
Db 497 GGAAGTCCAAATATGACTTATGACCCAGACGAATGATGACGCGCATGTGGCCTTTT 556  
QY 560 GGTTCCTAAGAACCGATTTAAGTGGACAGGTATGATGATCTTCTCAAGCAGAGT 619  
Db 557 GGTTCCTAAGAACCGATTTAAGTGGACAGGTATGATGATCTTCTCAAGCAGAGT 616  
QY 620 GGCAGTGAAGTGAACGAGGTGTTAAGGTTTAAATGAAGATATATACAGGCTCT 679  
Db 617 GGCAGTGAAGTGAACGAGGTGTTAAGGTTTAAATGAAGATATATACAGGCTCT 676

QY 680 GAAA 684  
Db 677 GAAA 681

RESULT 3  
CD557531  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
JOURNAL  
COMMENT

CD557531 896 bp mRNA linear EST 11-JUN-2003  
ABENCOUET 14413281 NIH MGC 180 Homo sapiens cDNA clone  
IMAGE:30350465 5', mRNA sequence.  
CD557531  
CD557531.1 GI:31583559  
EST.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.  
1 (bases 1 to 896)  
NIH-MGC http://mgi.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Daniela S. Gerhard, Ph.D.  
Office of Cancer Genomics  
National Cancer Institute / NIH  
Bldg. 31 Rm10A07 Bethesda, MD 20892  
Email: cgaabs-r@mail.nih.gov  
Tissue Procurement: Dr. Michael Brownstein  
cDNA Library Preparation: Invitrogen Corp  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: NDAM465 row: 3 column: 10  
High quality sequence stop: 592.  
Location/Qualifiers  
1. 896  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:30350465"  
/lab\_host="DH10B-Ton A ( T1 and T5 phage resistances )"  
/clone\_lib="NIH\_MGC\_180"  
/note="Organ: Testis; Vector: pCMV-Sport6.1; Site 1: NotI;  
Site 2: EcoRV (destroyed); Library is oligo-dT primed and  
directionally cloned (EcoRV site is destroyed upon  
cloning). Average insert size 1.68 kb. Library was  
constructed by (Invitrogen). Note: this is a NIH\_MGC  
Library."

ORIGIN  
Query Match 28.1%; Score 626; DB 14; Length 896;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 676; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 919 CATGCTAAAGCTGTTACTAGCTTACCTCTGCAAAATAACAGTATTCCTATCAT 978  
Db 1 CATGCTAAAGCTGTTACTAGCTTACCTCTGCAAAATAACAGTATTCCTATCAT 60

QY 979 ACTTGAGAGACGAGATTGATGCTTGCTCAACAGGCTCGGAGAGACGCGGCTT 1038  
Db 61 ACTTGAGAGACGAGATTGATGCTTGCTCAACAGGCTCGGAGAGACGCGGCTT 120

QY 1039 TCTCCTACCAATTTGGCTCATATGATGATGATGATGATGATGATGATGATGATGAT 1098  
Db 121 TCTCCTACCAATTTGGCTCATATGATGATGATGATGATGATGATGATGATGATGAT 180

QY 1099 GTTGCGAGAACGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1158  
Db 181 GTTGCGAGAACGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240

QY 1159 TTGGAAGCGAAGAAATTTCTTTGGAGACTGTGTAAGAGCTGTGTTATATATGAGGG 1218

Db 241 TTGGAAGCGAAGAAATTTCTTTGGAGACTGTGTAAGAGCTGTGTTATATATGAGGG 300

QY 1219 AACCCAGTGGGACATTCATATTCGACAAATAGTACAGGCTGTAATATATATGCTAC 1278  
Db 301 AACCCAGTGGGACATTCATATTCGACAAATAGTACAGGCTGTAATATATATGCTAC 360

QY 1279 TCTTGGAAGACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1338  
Db 361 TCTTGGAAGACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420

QY 1339 CTTAGTTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1398  
Db 421 CTTAGTTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480

QY 1399 GTTAAATTTCTTCCCGAAGATGCGATCAAGAGCAAGCCGCAACCTTATGTTCAAGTC 1458  
Db 481 GTTAAATTTCTTCCCGAAGATGCGATCAAGAGCAAGCCGCAACCTTATGTTCAAGTC 540

QY 1459 AACTTTTCCAGAGAAATTTAAAGTTGGTGTGCAAGATTTTAAAGCAATATATCTGT 1518  
Db 541 AACTTTTCCAGAGAAATTTAAAGTTGGTGTGCAAGATTTTAAAGCAATATATCTGT 600

QY 1519 TGTGCTGTTGACAAAGTGGTGGAGCATGTAGAGATGTTCAAGAGCCGTTCCAGAT 1578  
Db 601 TGTGCTGTTGACAAAGTGGTGGAGCATGTAGAGATGTTCAAGAGCCGTTCCAGAT 660

QY 1579 TGGCCAGTTTCAAAA 1595  
Db 661 TGGCCAGTTTCAAAA 677

RESULT 4  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
JOURNAL  
COMMENT

LOCUS  
602696209F1 NIH\_MGC\_97 Homo sapiens cDNA clone IMAGE:4828298 5',  
mRNA sequence.  
ACCESSION  
BG718234.1 GI:13997421  
BG718234  
EST.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.  
1 (bases 1 to 682)  
NIH-MGC http://mgi.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-r@mail.nih.gov  
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki  
Toshiyuki and Piero Carninci (RIKEN)  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LRAM10746 row: a column: 03  
High quality sequence stop: 682.  
Location/Qualifiers  
1. 682  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4828298"  
/lab\_host="DH10B"  
/clone\_lib="NIH\_MGC\_97"  
/note="Organ: testis; Vector: pBluescriptR (modified  
pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI  
(gcgag); Oligo-dT primed using primer  
5'-TTTTTTTTTTTTTNN-3', size-selected for average  
insert size 2.2 kb and normalized to ROT 5. This is a

primary library enriched for full-length clones and constructed using the Cap-trapper method (Garnicki, in preparation). Library constructed by M. Brownstein (NIH/NHRI, National Institutes of Health). Note: this is a NIH-MGC Library."

## ORIGIN

Query Match 27.7%; Score 616; DB 12; Length 682;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 616; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

20 GATGAGATGGGAGAGCAATCAACCTCATATGCTTCTATGTTCCCATTTGG 79  
67 GATGAGATGGGAGAGCAATCAACCTCATATGCTTCTATGTTCCCATTTGG 126  
80 AAGGATGATGATCTGGAGAAATGAGACATTTTACAGAGCTCAGCTCATGATCA 139  
127 AAGGATGATGATCTGGAGAAATGAGACATTTTACAGAGCTCAGCTCATGATCA 186  
140 GAAATGATGATGACCTTCTCGAAGATCATTTTCAAAAAGTGATTTGCTTGGG 199  
187 GAAATGATGATGACCTTCTCGAAGATCATTTTCAAAAAGTGATTTGCTTGGG 246  
200 CGGAATTTTGGAAACAGAGATGCTGAGTGAATTAAGAGATTAATCATCACAAG 259  
247 CGGAATTTTGGAAACAGAGATGCTGAGTGAATTAAGAGATTAATCATCACAAG 306  
260 GGTGCTTTTGGAGTGGAAAGATTTTGAACAGAGATTTTCAACAGCAGTTTGA 319  
307 GGTGCTTTTGGAGTGGAAAGATTTTGAACAGAGATTTTCAACAGCAGTTTGA 366  
320 GATGATGATGATCTGCTGTTTCTGGAGAGATCTATGATGATGCTGAGATTAATCCACA 379  
367 GATGATGATGATCTGCTGTTTCTGGAGAGATCTATGATGATGCTGAGATTAATCCACA 426  
380 CGGAATTTTGGAAACAGAGATGCTGAGTGAATTAAGAGATTAATCATCACAAG 439  
427 CGGAATTTTGGAAACAGAGATGCTGAGTGAATTAAGAGATTAATCATCACAAG 486  
440 GGGCCATACAGAGAGTGGAAAGTGAATTTTCCGAGTTTCCGTTGAGATTTGCTGA 499  
487 GGGCCATACAGAGAGTGGAAAGTGAATTTTCCGAGTTTCCGTTGAGATTTGCTGA 546  
500 GGAATGATGATGACCTTCTCGAAGATCATTTTCAAAAAGTGATTTGCTTGGG 559  
547 GGAATGATGATGACCTTCTCGAAGATCATTTTCAAAAAGTGATTTGCTTGGG 606  
560 GGTCTAGAGACAGATTTAAGTGAGCAGAGTGAATGATGATTTCTTCAAGAGAGT 619  
607 GGTCTAGAGACAGATTTAAGTGAGCAGAGTGAATGATGATTTCTTCAAGAGAGT 666  
620 GGCAGTGGAGTGAAC 635  
667 GGCAGTGGAGTGAAC 682

RESULT 5  
LOCUS B1560849 693 bp mRNA linear EST 05-SEP-2001  
DEFINITION 603254018F1 NIH\_MGC\_97 Homo sapiens cDNA clone IMAGE:529644 5',  
B1560849 mRNA sequence.  
ACCESSION B1560849  
VERSION B1560849.1 GI:15448163  
SOURCE EST.  
ORGANISM Homo sapiens (human)  
Homosapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 693)  
NIH-MGC http://mgi.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.

## FEATURES

Email: cgaabs@mail.nih.gov  
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
cDNA Library Preparation: Michael J. Brownstein (NHRI), Shiroki  
Tohiyuki and Piero Carninci (RIKEN)  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLU)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNLU at:  
http://image.llnl.gov  
Plate: LLM11750 row: c column: 05  
High quality sequence stop: 691.  
Location/Qualifiers

## source

1.693  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:529644"  
/lab\_host="DH10B"  
/clone\_id="NIH\_MGC\_97"  
/note="Organ: testis; Vector: pBluescriptR (modified  
pBluescript KS+); Site\_1: BamHI; Site\_2: SalI-XhoI  
(GCGAG); Oligo-dT primed using primer  
5'-TTTTTTTTTTTTTTVN-3', size-selected for average  
insert size 2.2 kb and normalized to 10<sup>5</sup>. This is a  
primary library enriched for full-length clones and  
constructed using the Cap-trapper method (Garnicki, in  
preparation). Library constructed by M. Brownstein  
(NIH/NHRI, National Institutes of Health). Note: this is  
a NIH-MGC Library."

## ORIGIN

Query Match 27.7%; Score 616; DB 12; Length 693;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 616; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

20 GATGAGATGGGAGAGCAATCAACCTCATATGCTTCTATGTTCCCATTTGG 79  
78 GATGAGATGGGAGAGCAATCAACCTCATATGCTTCTATGTTCCCATTTGG 137  
80 AAGGATGATGATCTGGAGAAATGAGACATTTTACAGAGCTCAGCTCATGATCA 139  
138 AAGGATGATGATCTGGAGAAATGAGACATTTTACAGAGCTCAGCTCATGATCA 197  
140 GAAATGATGATGACCTTCTCGAAGATCATTTTCAAAAAGTGATTTGCTTGGG 199  
198 GAAATGATGATGACCTTCTCGAAGATCATTTTCAAAAAGTGATTTGCTTGGG 257  
200 CGGAATTTTGGAAACAGAGATGCTGAGTGAATTAAGAGATTAATCATCACAAG 259  
258 CGGAATTTTGGAAACAGAGATGCTGAGTGAATTAAGAGATTAATCATCACAAG 317  
260 GGTGCTTTTGGAGTGGAAAGATTTTGAACAGAGATTTTCAACAGCAGTTTGA 319  
318 GGTGCTTTTGGAGTGGAAAGATTTTGAACAGAGATTTTCAACAGCAGTTTGA 377  
320 GATGATGATGATCTGCTGTTTCTGGAGAGATCTATGATGATGCTGAGATTAATCCACA 379  
378 GATGATGATGATCTGCTGTTTCTGGAGAGATCTATGATGATGCTGAGATTAATCCACA 437  
380 CGGAATTTTGGAAACAGAGATGCTGAGTGAATTAAGAGATTAATCATCACAAG 439  
438 CGGAATTTTGGAAACAGAGATGCTGAGTGAATTAAGAGATTAATCATCACAAG 497  
440 GGGCCATACAGAGAGTGGAAAGTGAATTTTCCGAGTTTCCGTTGAGATTTGCTGA 499  
498 GGGCCATACAGAGAGTGGAAAGTGAATTTTCCGAGTTTCCGTTGAGATTTGCTGA 557  
500 GGAATGATGATGATCTGGAGAAATGAGACATTTTACAGAGCTCAGCTCATGATCA 559  
558 GGAATGATGATGATCTGGAGAAATGAGACATTTTACAGAGCTCAGCTCATGATCA 617  
560 GGTCTAGAGACAGATTTAAGTGAGCAGAGTGAATGATGATTTCTTCAAGAGAGT 619

Db 618 GCTTCTAGAACCACTATTAGTGCACAGTATGATGATCTTCTCAACAGCAAGT 6177  
 QY 620 GGCAGTGGAAAGTGAAC 635  
 Db 678 GGCAGTGGAAAGTGAAC 693

RESULT 6  
 BG717110 798 bp mRNA linear EST 08-MAY-2001  
 LOCUS 602689191.F1 NIH\_MGC\_97 Homo sapiens cDNA clone IMAGE:4821503 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BG717110  
 VERSION BG717110.1 GI:13996297  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 798)  
 AUTHORS NIH-MGC http://img.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapsb-remail.nih.gov  
 Tissue Procurement: Miklos Palokovits, M.D., Ph.D.  
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki  
 Toshiyuki, and Piero Carninci (RIKEN)  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.lnl.gov  
 Plate: LHAM0728 row: e column: 24  
 High quality sequence stop: 770.  
 Location/Qualifiers  
 1..798  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4821503"  
 /lab\_host="DH108"  
 /clone\_lib="NIH\_MGC\_97"  
 /note="Organ: Testis; Vector: pBluescript (modified  
 pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI  
 (gtcgag); Oligo-dT primed using primer  
 5'-TTTTTTTTTTTTTNN-3', size-selected for average  
 insert size 2.2 kb and normalized to 80% 5. This is a  
 primary library enriched for full-length clones and  
 constructed using the Cap-trapper method (Carninci, in  
 preparation). Library constructed by M. Brownstein  
 (NHGRI, National Institutes of Health). Note: this is  
 a NIH\_MGC Library."

ORIGIN  
 Query Match 27.5%; Score 612; DB 12; Length 798;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 662; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 29 TGGGAGCAGAAATCAACCTCATATGCTTCTCATATGTCATATTTGAGAGGATAG 88  
 Db 66 TGGGAGCAGAAATCAACCTCATATGCTTCTCATATGTCATATTTGAGAGGATAG 125  
 QY 89 TTTTCTGAGAAATGAGACAAATTTTAAACAGACTCCAGCTTCTATCAGAAATGAT 148  
 Db 126 TATCTGAGAAATGAGACAAATTTTAAACAGACTCCAGCTTCTATCAGAAATGAT 185  
 QY 149 GATGACCTTCTGAGAGATCATTTCAATGAAAGTGATTTGCTCGGGCGGAAATTT 208  
 Db 186 GATGACCTTCTGAGAGATCATTTCAATGAAAGTGATTTGCTCGGGCGGAAATTT 245  
 QY 209 GAAAACAGATGCTGCTGAGTGATTAATGAGAGATTAATCAATCCAAATGGGTGTTT 268

Db 246 GAAAACAGATGCTGCTGAGTGATTAATGAGAGATTAATCAATCCAAATGGGTGTTT 305  
 QY 269 GAGTTGAAAAGAGTTTGGAAAAGAGTTTCAAACAGCAGGTTTGAAGATGTGAT 328  
 Db 306 GAGTTGAAAAGAGTTTGGAAAAGAGTTTCAAACAGCAGGTTTGAAGATGTGAT 365  
 QY 329 AGCTTGCTTTCTGAGAGAGTCTAGTAAATGACTCCGAAATATCCAAACGGAACGA 388  
 Db 366 AGCTTGCTTTCTGAGAGAGTCTAGTAAATGACTCCGAAATATCCAAACGGAACGA 425  
 QY 389 GGGTTTCCAGAGAGCGGCTATCGAGATGGAATATTCAGAGGTTCCAGGCCATAC 446  
 Db 426 GGGTTTCCAGAGAGCGGCTATCGAGATGGAATATTCAGAGGTTCCAGGCCATAC 485  
 QY 449 AGAAGGTGGAAGAGTATGTTCCGAGGTTCCGAGAGATTTGCTTGAAGATCCA 508  
 Db 486 AGAAGGTGGAAGAGTATGTTCCGAGGTTCCGAGAGATTTGCTTGAAGATCCA 545  
 QY 509 AATATGACTAGACCCAGACGATATGATGAGGCGCTGGTGGCTTTTGGTCTAGA 568  
 Db 546 AATATGACTAGACCCAGACGATATGATGAGGCGCTGGTGGCTTTTGGTCTAGA 605  
 QY 569 AGACCAATATTAAGTGGACAGTAAATGATGATCTTCTCAAACAGAGTGGCAGTGA 628  
 Db 606 AGACCAATATTAAGTGGACAGTAAATGATGATCTTCTCAAACAGAGTGGCAGTGA 665  
 QY 629 AGTAAAGAGGTGGTTCAAAAGTTTAAATGAAGATATTAACAGCTCTGAAAGAT 688  
 Db 666 AGTAAAGAGGTGGTTCAAAAGTTTAAATGAAGATATTAACAGCTCTGAAAGAT 725  
 QY 689 TCT 691  
 Db 726 TCT 726

RESULT 7  
 CD557160 885 bp mRNA linear EST 11-JUN-2003  
 LOCUS AGENCOURT 14413590 NIH\_MGC\_180 Homo sapiens cDNA clone  
 DEFINITION IMAGE:30387531 5', mRNA sequence.  
 ACCESSION CD557160  
 VERSION CD557160.1 GI:31583228  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 885)  
 AUTHORS NIH-MGC http://img.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Daniela S. Gerhard, Ph.D.  
 Office of Cancer Genomics  
 National Cancer Institute / NIH  
 Bldg. 31 Rm10A07 Bethesda, MD 20892  
 Email: cgapsb-remail.nih.gov  
 Tissue Procurement: Dr. Michael Brownstein  
 cDNA Library Preparation: Invitrogen Corp  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.lnl.gov  
 Plate: NDMA57 row: p column: 04  
 High quality sequence stop: 671.  
 Location/Qualifiers  
 1..885  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:30387531"  
 /lab\_host="DH10B-Ton A ( T1 and T5 phage resistances)"  
 /clone\_lib="NIH\_MGC\_180"

FEATURES  
 source



Db	Query Match	Score	DB	Length	EST
Db	304	5076	100.0%	792	363
Qy	330	5076	100.0%	792	363
Db	364	5076	100.0%	792	363
Qy	380	5076	100.0%	792	363
Db	424	5076	100.0%	792	363
Qy	440	5076	100.0%	792	363
Db	484	5076	100.0%	792	363
Qy	500	5076	100.0%	792	363
Db	544	5076	100.0%	792	363
Qy	560	5076	100.0%	792	363
Db	604	5076	100.0%	792	363
RESULT 10					
LOCUS	B0429126				
DEFINITION	AGNCOURT 7901787 NIH_MGC_82 Homo sapiens cDNA clone IMAGE:6104663				
ACCESSION	B0429126				
VERSION	B0429126.1				
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Emmalyota, Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	1 (bases 1 to 792)				
TITLE	NIH-MGC http://mgs.nci.nih.gov/.				
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)				
COMMENT	Unpublished (1999)				
	Contact: Robert Strausberg, Ph.D.				
	Email: cgapds-remail.nih.gov				
	Tissue Procurement: CLONTECH				
	cDNA Library Preparation: CLONTECH Laboratories, Inc.				
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNU)				
	DNA Sequencing by: Agencourt Bioscience Corporation				
	Clone distribution: MGC clone distribution information can be				
	found through the I.M.A.G.E. Consortium/LNU at:				
	http://image.llnl.gov				
	Plate: LINC2342 row: n column: 24				
	High quality sequence stop: 496.				
FEATURES					
source	Location/Qualifiers				
	1..792				
	/organism="Homo sapiens"				
	/mol_type="mRNA"				
	/db_xref="taxon:9606"				
	/clone="IMAGE:6104663"				
	/lab_host="DH10B (TI phage-resistant)"				
	/clone_id="NIH MGC 82"				
	/note="Organ: testis; Vector: pDNR-LIB (Clontech); Site 1: SfiI (ggccatcggcc); Site 2: SfiI (ggccatcggcc); 5' adaptor sequence: 5'-CACGGCCATTATGACC-3' and 3' adaptor sequence: 5'-ATTCTAGAGGCCGAGCGCGCGAGATG-dt(30)BN-3' (where B = A, C, G and N = A, C, G, or T). Average insert size 1.35 kb (range 0.9-4.0 kb). 14/15 clones contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."				
ORIGIN					
Query Match	24.1%	Score 535	DB 13	Length 792	
Best Local Similarity	100.0%	Pred No. 5e-274			

Matches 535; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 GATGAGATTGGGAGAGAGAAATCAACCTCATATGTCTTCTATGTCCATATTTGAG 79  
 Db 67 GATTAAGATTGGGAGAGAGAAATCAACCTCATATGTCTTCTATGTCCATATTTGAG 126  
 QY 80 AAGATATGTTTCTTGGAGAAATGAGACATTTTAACGAGCTCCAGCTTCATATCA 139  
 Db 127 AAGATATGTTTCTTGGAGAAATGAGACATTTTAACGAGCTCCAGCTTCATATCA 186  
 QY 140 GAAATGATGATGAGACCTTCTCGAAGAGATCATTTCAATGAAAGTGGATTGCTCTG 199  
 Db 187 GAAATGATGATGAGACCTTCTCGAAGAGATCATTTCAATGAAAGTGGATTGCTCTG 246  
 QY 200 CGGAATTTTGGAAACAGAGATGCTGTGAGTGTATTAAGAGATTAATACATCAAA 259  
 Db 247 CGGAATTTTGGAAACAGAGATGCTGTGAGTGTATTAAGAGATTAATACATCAAA 306  
 QY 260 GGTGCTTTTGGAGTTGGAAGAGTTTGGAAACAGAGTTTTCAAACAGCAGTTTGA 319  
 Db 307 GGTGCTTTTGGAGTTGGAAGAGTTTGGAAACAGAGTTTTCAAACAGCAGTTTGA 366  
 QY 320 GATGTGATAGCTCTGCTTTCTGAGAGAGCTTATGATATGACTGCGAAGTAAATCCA 379  
 Db 367 GATGTGATAGCTCTGCTTTCTGAGAGAGCTTATGATATGACTGCGAAGTAAATCCA 426  
 QY 380 CGGAAGAGAGGTTTTCAGAGAGGCGCTATCGAGTGAATTAATTCAGAAAGCTCA 439  
 Db 427 CGGAAGAGAGGTTTTCAGAGAGGCGCTATCGAGTGAATTAATTCAGAAAGCTCA 486  
 QY 440 GGGCCATACAGAGAGGTGAGAGAGTATTTCCAGAGTTCCCGTGGAGATTTGGTCA 499  
 Db 487 GGGCCATACAGAGAGGTGAGAGAGTATTTCCAGAGTTCCCGTGGAGATTTGGTCA 546  
 QY 500 GGAAGTCCAAATATGACTTATGACCCAGAGATGTTGAGAGGCACTGGTGGCC 554  
 Db 547 GGAAGTCCAAATATGACTTATGACCCAGAGATGTTGAGAGGCACTGGTGGCC 601

## RESULT 11

CD557137 914 bp mRNA linear EST 11-JUN-2003  
 LOCUS ABEENCOURT\_14413272 NIH\_MGC\_180 Homo sapiens cDNA clone  
 DEFINITION IMAGE:30390297 5', mRNA sequence.

ACCESSION CD557137  
 VERSION CD557137.1 GI:31583205  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 914)  
 NIH-MGC http://mgi.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 CONTACT: Daniela S. Gerhard, Ph.D.  
 Office of Cancer Genomics  
 National Cancer Institute / NIH  
 Bldg. 31 Rm10A07 Bethesda, MD 20892  
 Email: cgabbs-remail.nih.gov  
 Tissue Procurement: Dr. Michael Brownstein  
 CDNA Library Preparation: Invitrogen Corp  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LINL at:  
 http://image.llnl.gov  
 Plate: NDAM465 row: c column: 10  
 High quality sequence start: 42  
 High quality sequence stop: 582.  
 Location/Qualifiers  
 1..914  
 /organism="Homo sapiens"

## FEATURES

source  
 1..914  
 /organism="Homo sapiens"

## ORIGIN

Query Match 23.4%; Score 521; DB 14; Length 914;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-266; Indels 0; Gaps 0;  
 Matches 521; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

/mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:30390297"  
 /lab\_host="DH10B-Ton A (T1 and T5 phage resistances)"  
 /clone\_11b="NIH\_MGC\_180"  
 /note="Organ: Testis; Vector: pCMV-Sport6.1; Site: 1: NotI;  
 Site 2: EcoRV (destroyed); Library is oligo-dT primed and  
 directionally cloned (EcoRV site is destroyed upon  
 cloning). Average insert size 1.68 kb. Library was  
 constructed by (Invitrogen). Note: this is a NIH\_MGC  
 library."

QY 1159 TTGGAAGCCAGAAATTTCTTTGGAGCTGTGTAAAGCTTTGTTATATATGAGGG 1218  
 Db 105 TTGGAAGCCAGAAATTTCTTTGGAGCTGTGTAAAGCTTTGTTATATATGAGGG 164  
 QY 1219 AACCCAGCTGGGACATTTCAATTCGACAAATAGTACAAAGCTGTATATATATGCTAC 1278  
 Db 165 AACCCAGCTGGGACATTTCAATTCGACAAATAGTACAAAGCTGTATATATATGCTAC 224  
 QY 1279 TCTGGAAGACTGATGATATCATATAGCGAAGAAAGATTGGTCTCAACAGATCAATA 1338  
 Db 225 TCTGGAAGACTGATGATATCATATAGCGAAGAAAGATTGGTCTCAACAGATCAATA 284  
 QY 1339 CTTAAGTTTGGATGAAGCTGATGCTATGTGATATGCGTTTGGTCCAGAAATGAATA 1398  
 Db 285 CTTAAGTTTGGATGAAGCTGATGCTATGTGATATGCGTTTGGTCCAGAAATGAATA 344  
 QY 1399 GTTAATTTCTGCCAGAAATGCGCATCAAGAAAGAGGCCAACCCTTATGTCAGTGC 1458  
 Db 345 GTTAATTTCTGCCAGAAATGCGCATCAAGAAAGAGGCCAACCCTTATGTCAGTGC 404  
 QY 1459 AACTTTTCCAGAGAAATTCAAAGGTGGCTGACAGATTTTAAAGTCAAAATATCTGTT 1518  
 Db 405 AACTTTTCCAGAGAAATTCAAAGGTGGCTGACAGATTTTAAAGTCAAAATATCTGTT 464  
 QY 1519 TGTTCGTGTGGACAAAGTGGTGGAGCATGTATGAGATTTTGAAGCAACCTTCCCAAGT 1578  
 Db 465 TGTTCGTGTGGACAAAGTGGTGGAGCATGTATGAGATTTTGAAGCAACCTTCCCAAGT 524  
 QY 1579 TGGCCAGTTCTCAAAAAGAGAAAGAGCTGTGAATTTCTGGAAGACATAGGGAGTGAAG 1638  
 Db 525 TGGCCAGTTCTCAAAAAGAGAAAGAGCTGTGAATTTCTGGAAGACATAGGGAGTGAAG 584  
 QY 1639 AACTATGCTCTTGTGTTAACTTAAGAAAAAGCAGATTTTA 1679  
 Db 585 AACTATGCTCTTGTGTTAACTTAAGAAAAAGCAGATTTTA 625

## RESULT 12

BTG19449 747 bp mRNA linear EST 08-MAY-2001  
 LOCUS 602690172F1 NIH\_MGC\_97 Homo sapiens cDNA clone IMAGE:4822528 5',  
 DEFINITION mRNA sequence.

ACCESSION BTG19449  
 VERSION BTG19449.1 GI:13998636  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE NIH-MGC http://mgi.nci.nih.gov/  
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
 TITLE Unpublished (1999)  
 JOURNAL Contact: Robert Strausberg, Ph.D.  
 COMMENT Email: cgabbs-remail.nih.gov  
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.



CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN)  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
 Plate: LLM10730 row: P column: 17  
 High quality sequence stop: 734.

## FEATURES

source

1. 747  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4826190"  
 /lab\_host="DH10B"  
 /clone\_lib="NIH\_MGC\_97"  
 /note="Organ: testis; Vector: pBluescript (modified pBluescript KS+); Site\_1: BamHI; Site\_2: SalI-XhoI (gtagag); Oligo-dt primed using primer 5'-TTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.2 kb and normalized to R0T 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 22.9%; Score 509; DB 12; Length 747;  
 Best Local Similarity 99.8%; Pred. No. 4e-260; 0; Indels 1; Gaps 1;  
 Matches 623; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

20 GATGAAGATTGGAGACAGAAATCAACCTCATATCTCTCTTCCATATTTGAG 79  
 103 GATGAAGATTGGAGACAGAAATCAACCTCATATCTCTCTTCCATATTTGAG 162  
 80 AAGGATAGGATCTCTGAGAAATGAGACATTTTAAACAGACTCCAGCTTCAATCA 139  
 163 AAGGATAGGATCTCTGAGAAATGAGACATTTTAAACAGACTCCAGCTTCAATCA 222  
 140 GAATGATGATGACCTTCTCGAAGAGATCAATTCATGAAATGATTTGCTCTGGG 199  
 223 GAATGATGATGACCTTCTCGAAGAGATCAATTCATGAAATGATTTGCTCTGGG 282  
 200 CGGAATTTTGGAAACAGAGATGCTGTGATGATTAACGAGATTAATCAATCCCAATG 259  
 283 CGGAATTTTGGAAACAGAGATGCTGTGATGATTAACGAGATTAATCAATCCCAATG 342  
 260 GGTGCTTTGGAGTTGGAAGAGTTTGGAAACAGAGTTTTCACACAGAGTTTGA 319  
 343 GGTGCTTTGGAGTTGGAAGAGTTTGGAAACAGAGTTTTCACACAGAGTTTGA 402  
 320 GATGCTGATAGCTCTGCTTTCTGGAGAGCTTAATGATGATCTGGAGATATATCCAA 379  
 403 GATGCTGATAGCTCTGCTTTCTGGAGAGCTTAATGATGATCTGGAGATATATCCAA 462  
 380 CGGAAGAGAGGTTTCCAAAGAGAGGCGCTATCGAGATGAGAA-ATAATTCAGAGCTTC 438  
 463 CGGAAGAGAGGTTTCCAAAGAGAGGCGCTATCGAGATGAGAA-ATAATTCAGAGCTTC 522  
 439 AGGCGCATACAGAGAGGTGAGAGAGTATCCAGAGTTCCGTTGAGAGATTTGCT 498  
 523 AGGCGCATACAGAGAGGTGAGAGAGTATCCAGAGTTCCGTTGAGAGATTTGCT 582  
 499 AGGAATCCAAATTAAGATCTTGAAGAGAGAGTATCCAGAGTTCCGTTGAGAGATTTGCT 558  
 583 AGGAATCCAAATTAAGATCTTGAAGAGAGAGTATCCAGAGTTCCGTTGAGAGATTTGCT 642  
 559 TGGTTCTAGAGACAGATTTAAGTGGCAAGATATGTTGATTTCTCAAGAGAGAG 618  
 643 TGGTTCTAGAGACAGATTTAAGTGGCAAGATATGTTGATTTCTCAAGAGAGAG 702

QY 619 TGGCAGTGAAGTGAACGAGGTGTACAA 648  
 DB 703 TGGCAGTGAAGTGAACGAGGTGTACAA 732

## RESULT 13

LOCUS

DEFINITION

ACCESSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

657 bp mRNA linear EST 08-MAY-2001  
 602693829F1 NIH\_MGC\_97 Homo sapiens CDNA clone IMAGE:4826190 5',  
 mRNA sequence.  
 BG177812  
 BG177812.1 GI:13996999  
 EST.  
 Homo sapiens (human)  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NIH-MGC <http://mgs.nci.nih.gov/>.  
 1 (bases 1 to 657)  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
 Tissue Procurement: Miklos Palokovits, M.D., Ph.D.  
 CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN)  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
 Plate: LLM10740 row: i column: 07  
 High quality sequence stop: 655.  
 Location/Qualifiers  
 1. 657  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4826190"  
 /lab\_host="DH10B"  
 /clone\_lib="NIH\_MGC\_97"  
 /note="Organ: testis; Vector: pBluescript (modified pBluescript KS+); Site\_1: BamHI; Site\_2: SalI-XhoI (gtagag); Oligo-dt primed using primer 5'-TTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.2 kb and normalized to R0T 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: this is a NIH\_MGC Library."

## FEATURES

source

20 GATGAAGATTGGAGACAGAAATCAACCTCATATGCTCTCTTCCATATTTGAG 79  
 57 GATGAAGATTGGAGACAGAAATCAACCTCATATGCTCTCTTCCATATTTGAG 116  
 80 AAGATAGGATTTCTGAGAAATGAGACATTTTAAACAGAGCTCCAGCTTCAATCA 139  
 117 AAGATAGGATTTCTGAGAAATGAGACATTTTAAACAGAGCTCCAGCTTCAATCA 176  
 140 GAATGATGATGACCTTCTCGAAGAGATCAATTCATGAAATGATTTGCTCTGGG 199  
 177 GAATGATGATGACCTTCTCGAAGAGATCAATTCATGAAATGATTTGCTCTGGG 236  
 200 CGGAATTTTGGAAACAGAGATGCTGTGATGATTAACGAGATTAATCAATCCCAATG 259  
 237 CGGAATTTTGGAAACAGAGATGCTGTGATGATTAACGAGATTAATCAATCCCAATG 296

## ORIGIN

Query Match 21.6%; Score 480; DB 12; Length 657;  
 Best Local Similarity 99.8%; Pred. No. 1.3e-244;  
 Matches 600; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 20 GATGAAGATTGGAGACAGAAATCAACCTCATATGCTCTCTTCCATATTTGAG 79  
 DB 57 GATGAAGATTGGAGACAGAAATCAACCTCATATGCTCTCTTCCATATTTGAG 116  
 QY 80 AAGATAGGATTTCTGAGAAATGAGACATTTTAAACAGAGCTCCAGCTTCAATCA 139  
 DB 117 AAGATAGGATTTCTGAGAAATGAGACATTTTAAACAGAGCTCCAGCTTCAATCA 176  
 QY 140 GAATGATGATGACCTTCTCGAAGAGATCAATTCATGAAATGATTTGCTCTGGG 199  
 DB 177 GAATGATGATGACCTTCTCGAAGAGATCAATTCATGAAATGATTTGCTCTGGG 236  
 QY 200 CGGAATTTTGGAAACAGAGATGCTGTGATGATTAACGAGATTAATCAATCCCAATG 259  
 DB 237 CGGAATTTTGGAAACAGAGATGCTGTGATGATTAACGAGATTAATCAATCCCAATG 296

OY	260	GGTGGTTTGGAGTTGGAAAGAGTTTGGAAACAGAGTTTTCAAACAGCAGGTTTGA	319
Db	297	GGTGGTTTGGAGTTGGAAAGAGTTTGGAAACAGAGTTTTCAAACAGCAGGTTTGA	356
OY	320	GATGCTGATAGCTCTGTTTCTGGAGAGTCTAGTAA TGACTGCGAAGATATCCAA	379
Db	357	GATGCTGATAGCTCTGTTTCTGGAGAGTCTAGTAA TGACTGCGAAGATATCCAA	416
OY	380	CGAACAAGGCTTTTCCAAAGAGGCGGCTTTCAGATGGAATAATTCA GAAGCTTC	438
Db	417	CGAACAAGGCTTTTCCAAAGAGGCGGCTTTCAGATGGAATAATTCA GGAAGCTTC	476
OY	439	AGGGCCATACAGAAAGGTCGAAGAGTAGTTTCGAGAGTTCGCTGGAGGATTTGGTCT	498
Db	477	AGGGCCATACAGAAAGGTCGAAGAGTAGTTTCGAGAGTTCGCTGGAGGATTTGGTCT	536
OY	499	AGGAAGTCCAAATTAATGAATTAGACCCAGACGAAATGATGCAAGCCACTGTGGGCTTTT	558
Db	537	AGGAAGTCCAAATTAATGAATTAGACCCAGACGAAATGATGCAAGCCACTGTGGGCTTTT	596
OY	559	TGGTTCTTGAAAGACCAAGTATTTAGTGGCAACAGTATGCTGATTACTTCTCAAGACAGAA	618
Db	597	TGGTTCTTGAAAGACCAAGTATTTAGTGGCAACAGTATGCTGATTACTTCTCAAGACAGAA	656
OY	619	T 619	
Db	657	T 657	

RESULT 14  
CD557866  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

933 bp mRNA linear EST 11-JUN-2001  
CD557866 AGSCOURT 14423254 NIH MGC\_180 Homo sapiens cDNA clone  
IMAGE:30350020 5', mRNA sequence.  
CD557866  
CD557866.1 GI:31583934  
EST.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 933)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Daniela S. Gerhard, Ph.D.  
Office of Cancer Genomics  
National Cancer Institute / NIH  
Bldg. 31 Rm10A07 Bethesda, MD 20892  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: Dr. Michael Brownstein  
cDNA Library Preparation: Invitrogen Corp  
cDNA Library Arrayed By: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: NDAM464 row: g column: 21  
High quality sequence start: 19  
High quality sequence stop: 592.

FEATURES

source

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30390020"
/lab_host="DH10B-Ton A (T1 and T5 phage resistances)"
/clone_lib="N1H_MGC_180"
/name="Organ: Testis; Vector: pCMV-Sport6.1; Site_1: NotI
site_2: EcoRV (destroyed); Library is ciso-dt primed and
directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.68 Kb. Library was
constructed by Invitrogen). Note: this is a N1H_MGC

```

ORIGIN	Library."
Query Match	21.2%; Score 471; DB 14; Length 933;
Best Local Similarity	100.0%; Pred. No. 8.6e-240;
Matches 471; Conservative	0; Mismatches 0; Indels 0; Gaps 0.

Qy	1012	AACGGGGCTGGGAAAGCTGGGCGCTTTTCTCTCCAAATTTGGCTCATGATGATCA	1071
Db	51	AACGGGGCTGGGAAAGCTGGGCGCTTTTCTCTCCAAATTTGGCTCATGATGATCA	110
Qy	1072	TGGAATACTGCCAGTGGTTTTAAAGATGCGAGAACCAAGATATATTGTAGCAC	113
Db	111	TGGAAATACTGCCAGTGGTTTTAAAGATGCGAGAACCAAGATATATTGTAGCAC	170
Qy	1132	AACTCGAATTTGGTCAACACGATTTATTTGGAAGCCGAAATTTCTTTGGGACTTG	1191
Db	171	AACTCGAATTTGGTCAACACGATTTATTTGGAAGCCGAAATTTCTTTGGGACTTG	230
Qy	1192	TGTAAAGCTGTGTATATATATGCGGGAAACCAGCTGGGACATTCATTCGAAATAGT	1251
Db	231	TGTAAAGCTGTGTATATATATGCGGGAAACCAGCTGGGACATTCATTCGAAATAGT	290
Qy	1252	ACAAAGCTGTATATATTATATGTGCTACCTCTGGAAAGATGATGATATCATATGGAAGA	1311
Db	291	ACAAAGCTGTATATATTATGTGCTACCTCTGGAAAGATGATGATATCATATGGAAGA	350
Qy	1312	AAAGATTGCTCAACACGATCAATACTTATGTTTGGATGAAGCTGATCGCATGWTGGA	1371
Db	351	AAAGATTGCTCAACACGATCAATACTTATGTTTGGATGAAGCTGATCGCATGWTGGA	410
Qy	1372	TATGGGTTTTGGTCCAGAAATGAAGAAATTAATTTCTTGCCAGGAATGCCATCAAAGA	1431
Db	411	TATGGGTTTTGGTCCAGAAATGAAGAAATTAATTTCTTGCCAGGAATGCCATCAAAGA	470
Qy	1432	ACAAGGCCAAACCCCTTATGTCACATGCAACTTTTCCAGAGAAATTCAAAG 1482	
Db	471	ACAAGGCCAAACCCCTTATGTCACATGCAACTTTTCCAGAGAAATTCAAAG 521	

RESULT	15
LOCUS	Bg717354
DEFINITION	Bg717354 651 bp mRNA linear EST 08-MAY-2003 602698688B1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4822149 5' , mRNA sequence.
ACCESSION	Bg717354
VERSION	Bg717354
KEYWORDS	Bg717354.1 GI:13996541
SOURCE	EST .
ORGANISM	Homo sapiens (human)
AUTHORS	Homo sapiens
TITLE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 651) NIH-MGC http://mgc.nci.nih.gov/. National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Straussberg, Ph.D.
JOURNAL COMMENT	

FEATURES  
source  
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki  
Toshiyuki and Pietro Carninci (RIKEN)  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: WGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.lnl.gov>  
plate: LRAM10729 row: p column: 22  
High quality sequence stop: 651.  
Location/Qualifiers  
1..651  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"

```
/clone="IMAGE:4822149"  
/lab host="DH10B"  
/clone lib="NIH_MGC_97"  
/note="Organ: testis; Vector: pBluescriptR (modified  
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI  
(GTCGAG); Oligo-dT primed using primer  
5'-TTTTTTTTTTTTTTVN-3', size-selected for average  
insert size 2.2 kb and normalized to R01 5. This is a  
primary library enriched for full-length clones and  
constructed using the Cap-trapper method (Carninci, in  
preparation). Library constructed by M. Brownstein  
(NIH/NHGRI, National Institutes of Health). Note: this is  
a NIH_MGC Library."
```

## ORIGIN

```
Query Match      21.1%; Score 469; DB 12; Length 651;  
Best Local Similarity 100.0%; Pred. No. 9,6e-239;  
Matches 469; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 29 TGGGAAGCAGAAATCAACCTCATATGCTCTCTATGTTCCCATATTGAGAAAGATAG 88  
Db 77 TGGGAAGCAGAAATCAACCTCATATGCTCTCTATGTTCCCATATTGAGAAAGATAG 136  
  
QY 89 TATTCGAGAAATGAGACATTTTAAACAGACTCCAGCTTCATCATCAGAAATGAT 148  
Db 137 TATTCGAGAAATGAGACATTTTAAACAGACTCCAGCTTCATCATCAGAAATGAT 196  
  
QY 149 GATGACCTTCTGAAAGATCATTTTCAAGAAAGTGATTTGCTCTGCGCGAAATTTT 208  
Db 197 GATGACCTTCTGAAAGATCATTTTCAAGAAAGTGATTTGCTCTGCGCGAAATTTT 256  
  
QY 209 GAAACAGAGATGCTGAGTGTATTAAGCGAGATTAATACCAATGGATGGTGT 268  
Db 257 GAAACAGAGATGCTGAGTGTATTAAGCGAGATTAATACCAATGGATGGTGT 316  
  
QY 269 GAGATTGAAAGATTTTGGAAACAGAGTTTTCAAACAGCAGTTTGAAGATGAT 328  
Db 317 GAGATTGAAAGATTTTGGAAACAGAGTTTTCAAACAGCAGTTTGAAGATGAT 376  
  
QY 329 AGCTCTGCTTCTGAGAGAGTCTAGTAAATGACTGGAGAGATTAATCCAAACAGAAACAGA 388  
Db 377 AGCTCTGCTTCTGAGAGAGTCTAGTAAATGACTGGAGAGATTAATCCAAACAGAAACAGA 436  
  
QY 389 GGGTTTTCCAAAGAGCGGCTATCGAGATGAAATTAATTCAGAACTTCAGGCGCATAC 448  
Db 437 GGGTTTTCCAAAGAGCGGCTATCGAGATGAAATTAATTCAGAACTTCAGGCGCATAC 496  
  
QY 449 AGAAGAGGTGAAAGAGTACTTCCAGAGTTGCCGTGAGAGATTTGATC 497  
Db 497 AGAAGAGGTGAAAGAGTACTTCCAGAGTTGCCGTGAGAGATTTGATC 545
```

Search completed: July 27, 2004, 18:08:12  
Job time : 5702 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 27, 2004, 03:14:19 ; Search time 5696 Seconds  
(without alignments)  
11659.672 Million cell updates/sec

Title: US-09-714-865B-1

Perfect score: 2224

Sequence: 1 actgtgaagtcacatgsggg999.....aagtcgtgttgcagca 2224

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estdb:\*  
2: em\_estdb:\*  
3: em\_estdb:\*  
4: em\_estdb:\*  
5: em\_estdb:\*  
6: em\_estdb:\*  
7: em\_estdb:\*  
8: em\_estdb:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_est3:\*  
12: gb\_est3:\*  
13: gb\_est3:\*  
14: gb\_est3:\*  
15: em\_estdb:\*  
16: em\_estdb:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_hum:\*  
20: em\_gss\_hum:\*  
21: em\_gss\_hum:\*  
22: em\_gss\_hum:\*  
23: em\_gss\_hum:\*  
24: em\_gss\_hum:\*  
25: em\_gss\_hum:\*  
26: em\_gss\_hum:\*  
27: em\_gss\_hum:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2200.8	99.0	2523	11	BC030638 Homo sapi
2	1539.8	63.2	2755	11	AK014844 Mus muscu
3	772.2	34.7	896	14	CD557531 AGENCOURT
4	741.8	33.4	850	14	CD251828 AGENCOURT

Result No.	Score	Query Match	Length	ID	Description
5	733.8	33.0	933	14	CD557866 AGENCOURT
6	729.6	31.2	885	14	CD557160 AGENCOURT
7	694.4	32.8	798	12	BG717110 AGENCOURT
8	652.2	29.3	822	12	B1464267 AGENCOURT
9	649.8	29.2	914	14	CD557137 AGENCOURT
10	643.2	28.9	792	13	BQ429126 AGENCOURT
11	636.2	28.6	747	12	BG719449 AGENCOURT
12	631.8	28.4	682	12	BG718234 AGENCOURT
13	631.8	28.4	693	12	B1560849 AGENCOURT
14	631.2	28.4	868	14	CD358073 AGENCOURT
15	627.8	28.2	830	12	B1462207 AGENCOURT
16	623.8	28.0	918	12	B1561001 AGENCOURT
17	621.8	28.0	846	12	BG718634 AGENCOURT
18	617.6	27.8	981	14	CD359812 AGENCOURT
19	608.8	27.4	699	12	BG717822 AGENCOURT
20	608.8	27.3	758	12	B1559479 AGENCOURT
21	607.8	27.3	939	14	CD300339 AGENCOURT
22	606.2	27.3	861	12	B1461393 AGENCOURT
23	604.8	27.2	657	12	BG717812 AGENCOURT
24	596.8	26.8	743	12	B1463502 AGENCOURT
25	593.2	26.7	884	12	B1560499 AGENCOURT
26	591.8	26.6	823	13	B0939339 AGENCOURT
27	577.4	26.0	805	14	CB230120 AGENCOURT
28	575.8	25.9	651	12	BG717354 AGENCOURT
29	571.2	25.7	700	12	BG718861 AGENCOURT
30	555.8	25.0	847	12	B1562624 AGENCOURT
31	554.8	25.0	810	12	B1463846 AGENCOURT
32	536.2	24.1	841	12	B1459115 AGENCOURT
33	515.4	23.2	917	12	B1463447 AGENCOURT
34	483.6	21.7	546	12	BMS39751 AGENCOURT
35	482.6	21.7	561	14	CD561008 AGENCOURT
36	469.4	21.1	690	14	CD767258 AGENCOURT
37	449.8	20.2	670	13	BQ443738 AGENCOURT
38	445.2	20.0	482	9	AL042306 DXF2p434M
39	420.2	18.9	780	10	AM153705 F124905.Y
40	405.8	18.2	759	14	CK026554 AGENCOURT
41	393.2	17.7	635	13	BY716292 AGENCOURT
42	382.8	17.6	512	28	CC183462 XG823 Bay
43	389.8	17.5	726	10	BR611759 de88d02.Y
44	380.2	17.1	604	12	B1460628 AGENCOURT
45	371.8	16.7	4617	11	BC007668 Homo sapi

## ALIGNMENTS

RESULT 1  
BC030638  
LOCUS  
DEFINITION  
Homo sapiens CDNA clone IMAGE:4822528, containing frame-shift errors  
ACCESSION  
BC030638  
VERSION  
BC030638.1 GI:21040391  
KEYWORDS  
Homo sapiens (human)  
SOURCE  
Homo sapiens  
ORGANISM  
Homo sapiens  
REFERENCE  
1 (bases 1 to 2523)  
Serausberg R.L., Feingold R.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buelow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C., Kane S.S., Loguercio N.A., Peters G.J., Abramson R.D., Mullany S.J., Bosak S.A., McEwan P.J., McKernan K.J., Hale S., Garcia A.M., Gay L.J., Huijck S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Rahney J., Helton E., Ketterman M., Madan A., Rodrigues S., Sanchez A., Whitting W., Madan A., Young A.C., Shevchenko Y.,

Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
Butlerfield, Y.S., Krzywicki, M.I., Skalska, U., Smalusz, D.E.,  
Scherer, A., Schein, J.E., Jones, S.D., and Marra, M.A.  
Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Strausberg, R.  
Direct Submission  
Submitted (20-MAY-2002) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk

Email: [gcaps-remail.nih.gov](mailto:gcaps-remail.nih.gov)  
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki  
Toshiyuki and Piero Carninci (RIKEN)  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Institute for Systems Biology  
<http://www.systemsbio.org>  
Contact: [amadan@systemsbiology.org](mailto:amadan@systemsbiology.org)  
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha  
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
Series: IRAK Plate: 33 Row: d Column: 8  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 9507236  
This clone has the following problem: frame shifted.

FEATURES  
Source

1..2523  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGH:4822528"  
/issue\_type="Testis"  
/clone\_lib="NIH MGC\_97"  
/lab\_host="DH10B"  
/note="Vector: pBluescript"

ORIGIN

Query Match 99.0%; Score 2200.8; DB 11; Length 2523;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 2216; Conservative 0; Mismatches 7; Indels 1; Gaps 1;  
QY 1 ACTGAGTCACCATGGGGATGAGATTGGGAGCAGAAATCAACCTCATATGCTTC 60  
DB 85 ACTTGAGCCACCATGGGATGAGATTGGGAGCAGAAATCAACCTCATATGCTTC 144  
QY 61 CTATGTTCCCATATTTGAGAGAGATAGTATTCTGGAGAAATGAGACAATTTTAACG 120  
DB 145 CTATGTTCCCATATTTGAGAGAGATAGTATTCTGGAGAAATGAGACAATTTTAACG 204  
QY 121 GACTCCAGCTTCATCATCAAAATGATGATGACCTTCGAGAGATCATTTTCATGAA 180  
DB 205 GACTCCAGCTTCATCATCAAAATGATGATGACCTTCGAGAGATCATTTTCATGAA 264  
QY 181 AAGTGATTTGCTCTGGGCGGAATTTTGGAAAACAGATGCTGGTGAAGTAAATACG 240  
DB 265 AAGTGATTTGCTCTGGGCGGAATTTTGGAAAACAGATGCTGGTGAAGTAAATACG 324  
QY 241 AATATATACATCAACATGGGTGTTTGGAGTTGAAAAGTTTGGAAAACAGAGTTT 300  
DB 325 AATATATACATCAACATGGGTGTTTGGAGTTGAAAAGTTTGGAAAACAGAGTTT 384  
QY 301 TTCAAAACGAGGTTTGAAGATGATGATGCTTCGAGAGAGTCTAGTAATGA 360  
DB 385 TTCAAAACGAGGTTTGAAGATGATGATGCTTCGAGAGAGTCTAGTAATGA 444

QY 361 CTGCAGATATATCCAAACGGAACAGAGGTTTTCAGAGAGCGGCTATCGAGATGG 420  
DB 445 CTGCAGATATATCCAAACGGAACAGAGGTTTTCAGAGAGCGGCTATCGAGATGG 504  
QY 421 AATATATCAAGGTTTGAAGGCTATCAGAGAGGTTGGAAGAGTGTGTTCCAGGTTG 480  
DB 505 AATATATCAAGGTTTGAAGGCTATCAGAGAGGTTGGAAGAGTGTGTTCCAGGTTG 564  
QY 481 CCGTGAGGATTTGGCTTGAAGAGTCCAAATATGACTTGAACCCAGACGATATATGA 540  
DB 565 CCGTGAGGATTTGGCTTGAAGAGTCCAAATATGACTTGAACCCAGACGATATATGA 624  
QY 541 GCGCATGTTGGCTTTTGGTTTGAAGACCGATTTTAAGTGGACAGGTAATGGTGA 600  
DB 625 GCGCATGTTGGCTTTTGGTTTGAAGACCGATTTTAAGTGGACAGGTAATGGTGA 684  
QY 601 TACTTCCAAAGCAGAGTGGCAGTGAAGTGAACGAGTGGTTTCAAGGTTTAAATGA 660  
DB 685 TACTTCCAAAGCAGAGTGGCAGTGAAGTGAACGAGTGGTTTCAAGGTTTAAATGA 744  
QY 661 AGAAGTATTAACAGGCTCTGGAAAAGATTTCTGGAAATCAAGACAGAGAGAGAAAG 720  
DB 745 AGAAGTATTAACAGGCTCTGGAAAAGATTTCTGGAAATCAAGACAGAGAGAGAAAG 804  
QY 721 TAGTGATATCAAGACCAAAAGTGAACCTTCAATCCCTCTCCACCTGAGATAGGA 780  
DB 805 TAGTGATATCAAGACCAAAAGTGAACCTTCAATCCCTCTCCACCTGAGATAGGA 864  
QY 781 CTCCATCTTTGACATTTATGACAGAGCATTAACCTTGACAAATACGACATATTTCTGT 840  
DB 865 CTCCATCTTTGACATTTATGACAGAGCATTAACCTTGACAAATACGACATATTTCTGT 924  
QY 841 GGAAGTGTGAGCATATGACACACAGAGATTTGACCTTTGAGAAAGCTATATCTCG 900  
DB 925 GGAAGTGTGAGCATATGACACACAGAGATTTGACCTTTGAGAAAGCTATATCTCG 984  
QY 901 TCAGACATGAATTAACACATTTGCTAAAGTGGTTTACTTAACCTTCTCTGCAAAA 960  
DB 985 TCAGACATGAATTAACACATTTGCTAAAGTGGTTTACTTAACCTTCTCTGCAAAA 1044  
QY 961 ATACAGATTTCTATCATCTTTGACAGAGAGATTTGATGCTTTGCTCAAAACAGGTC 1020  
DB 1045 ATACAGATTTCTATCATCTTTGACAGAGAGATTTGATGCTTTGCTCAAAACAGGTC 1104  
QY 1021 TGGGAAGACTGGGCTTTTCTCTCAATTTTGGCTCATATGATGATGATGATGATTAAC 1080  
DB 1105 TGGGAAGACTGGGCTTTTCTCTCAATTTTGGCTCATATGATGATGATGATGATTAAC 1164  
QY 1081 TGCAGTCTGTTTTAAAGAGTTGACAGAACCAAGTATTTATTTGACCACTGACAG 1140  
DB 1165 TGCAGTCTGTTTTAAAGAGTTGACAGAACCAAGTATTTATTTGACCACTGACAG 1224  
QY 1141 ATTGATCAACAGATTTATTTGAGAGCAGAAAATTTCTTTTGGACTTGTGTAAGAGC 1200  
DB 1225 ATTGATCAACAGATTTATTTGAGAGCAGAAAATTTCTTTTGGACTTGTGTAAGAGC 1284  
QY 1201 TGTGTTATATATGAGGAGAAACCAAGCTGGGACATTCATTTGACAAATATGTAACAAGCTG 1260  
DB 1285 TGTGTTATATATGAGGAGAAACCAAGCTGGGACATTCATTTGACAAATATGTAACAAGCTG 1344  
QY 1261 TAATATATATATGAGTACCTCTGAGAGATGATGATATCATATGAGGAGAAAGTATG 1320  
DB 1345 TAATATATATATGAGTACCTCTGAGAGATGATGATATCATATGAGGAGAAAGTATG 1403  
QY 1321 TCTCAAAACAGATCAATACCTTATTTGATGAGAGCTGATCCAGATGATGATGATG 1380  
DB 1404 TCTCAAAACAGATCAATACCTTATTTGATGAGAGCTGATCCAGATGATGATGATG 1463  
QY 1381 TGGTCCAGAAATGAAGATTTATTTCTTGGCCAGAAATGCGCATCAAGAGAACAGGCGCA 1440  
DB 1464 TGGTCCAGAAATGAAGATTTATTTCTTGGCCAGAAATGCGCATCAAGAGAACAGGCGCA 1523



FEATURES  
SOURCE

CDS

Location/Qualifiers  
1..2759  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM\_DB:4921508G22"  
/db\_xref="MGI:1910815"  
/db\_xref="taxon:10090"  
/clone="4921508G22"  
/sex="male"  
/tissue\_type="testis"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="adult"  
42..2150  
/note="unnamed protein product; DEAD (aspartate-glutamate-alanine-aspartate) box polypeptide 4 (MGI:MGI:102670)  
putative"  
/codon\_start=1  
/protein\_id="BAB29578.1"  
/db\_xref="GI:12852922"  
/translation="MGDEDEWEATILKPHVSSYVPEKDKYSSGANGDTFNRSSASD  
IGSSKKENSTTGGRGKRGKGNRGRFLNKKPEEGSSGFWKESNNDEEDNTRRGF  
SKRGCGODNDSPASPPRRGGRGSPFCRGFGLRFRVSDQDGTORGGLGSR  
KPAASDGNEDYOSRSRSGRGKGLNEVVTGSKNWKSETEGESDSQGRVT  
YLPPEDEDSIFAHYQTGINFDKYDTILVSGHDPAITLFEANLCOTLNNTI  
AKAGYKTLIPVKYSIPVLGRDLMACTQSGKTAFLILAHMRDGTIYASRF  
ELQEPICIVAPRELINOYLEAKFSFGTCVRAVYVIGTQFGHSYROIYQGNIL  
CATGRIMDIKEXRIGLKVYLVDEADRMIDMGFGEKLLISCGMPKRGROT  
LFSAPREIOLAGDPLKSSYLPAVONGGACRDVQITLLQYQSKREKVEIL  
RNIGDERIVFYETKKADPIATFLQEKISTLSHGBRERERQALSFEDTDNHL  
LVATSVARGLDLENVQVINFPLPSTIDYVHRIGRTGRCNTGRLASFEDTDNHL  
LAQPLKVLSDAQDVPALWLEIAFSTVPSFSSSTRGAVFASVDTRKMYQGHLL  
NTGSSSQAPNVDDESMD"  
2734..2739  
polya\_signal  
/note="putative"  
2759  
polya\_site  
/note="putative"

## ORIGIN

Query Match 69.2%; Score 1539.8; DB 11; Length 2759;  
Best Local Similarity 86.1%; Pred. No. 0;  
Matches 1770; Conservative 0; Mismatches 262; Indels 23; Gaps 5;  
QY 181 AAGTGAATTTGCTCTGGCGGGAATTTGAAAACAGAGTCTGTGATGTAATAGCG 240  
DB 134 AATGAGACACTTTTACAGACCTTCACTTCACTAGATATTGGCGAGTCAATAAAA 193  
QY 241 AGATAATACATCCAAATGGTGTGTTTGAAGTGAAGAAGTTTGGAAACAGAGTTT 300  
DB 194 AGGAACACATCTACAACTGTGCTTTGGAAGAGAAAGGCTTTGGAACAGAGTTT 253  
QY 301 TTCAAACAGAGTTTGAAGTGTGATAGCTCTGCTTTCTGAGAGAGTCTAGTATGA 360  
DB 254 TTAAATATCAAGTTTGAAGAGTGTAGCTCTGCTTTCTGGAAGAAGTCTAATATGA 313  
QY 361 CTGCGAAGATATCCAAACAGAGAGAGGTTTCCAAAGAGGCGGCTATCCAGATG 420  
DB 314 CTGTGAAGATATCAAGCTCGAAGAGAGGTTTTCAAAGAGGCTGCTGCAAGATG 373  
QY 421 AATAATTCAGAGCTTCAAGGCAATACAGAGAGTGAAGAGGTAAGTTTCCAGAGTTG 480  
DB 374 AAACATTCAGAGCATCAGGCCCGTTCAGAGAGAGGAGAGAGCACTTTCCAGAGCTG 433  
QY 481 CCGTGAAGATTTGGTCTAGAGAGTCCAAATATGACTTAAACCGACGATGTATGCA 540  
DB 434 CCGTGAAGATTTGGTCTAGAGAGCAAAATAGTAAATCTGACCAAGATCAGGGAGACA 493  
QY 541 GCGCACTGGTGGCTTTTGGTCTAGAGAGCACTAATTAAGTGCACAGGTAATGTGA 600  
DB 494 GCGTGGTGGCTTTTGGTCTAGAGAGCAACGACGACGATGATTAAGGCAATGTGTA 553  
QY 601 TACTTCTCAAGCAGAGTGCAGTGAAGTGAACGAGGTGTTTCAAAAGGTTTAAATGA 660

DB 554 CACTTACCAAGACGAG-----AGTGAAGTGTGAGGTGTTTCAAAAGTTTAAATGA 607  
QY 661 AGAATTAATTAACAGGCTCTGGAAAAGAAATTTGGAAATCAGAAACAGAGAGAGAAAG 720  
DB 608 AGAAGTAAATACAGGTTCTGGAAAAGAAATTTGGAAATCAGAAATCAGAGAGTGAAG 667  
QY 721 TAGGATATCTAAGAGACCAAAAGTACATATCCCCCTCTCTCACTTGAAGATGAGGA 780  
DB 668 CAGGATATGTAAGGTCCAAAAGTAAATATATCCCCCTCTCTCACTTGAAGATGAGGA 727  
QY 781 CTCCATCTTTCAGATTAATCAGAGGCAATTAATTCGACAAATACACACTATTTCTGT 840  
DB 728 CTCCATCTTTCAGATTAATCAGAGGCAATTAATTTGAATTAATACACACTATTTCTGT 787  
QY 841 CGAAGTGTCTGGAATGATGACCAACAGCAATTTCTGACTTTGAAGAGCTAATCTGTG 900  
DB 788 TGAAGTATCTGGAATGATGACCAACAGCAATTTTGAATTAATTAATCTATCTGTG 847  
QY 901 TCAGACCTGAATTAACAACTTGTAAAGTGGTAACTAAGTTTCTCTGTGCAAAA 960  
DB 848 TCAGACCTGAATTAACAACTTGTAAAGTGGTAACTAAGTTTCTCTGTGCAAAA 907  
QY 961 ATACAGATTTCTATATCTTGTGAGAGCAGATTTGTATGCTGTGCAAAACAGGCTC 1020  
DB 908 GTACAGATTTCTATATCTTGTGAGAGCAGATTTGTATGCTGTGCAAAACAGGCTC 967  
QY 1021 TGGGAAGACGCGGCTTTCTCTCAATTTTGGCTCATATGATGATGATGATGATGATGAT 1080  
DB 968 TGGGAAGACGCGGCTTTCTCTCAATTTTGGCTCATATGATGATGATGATGATGATGAT 1027  
QY 1081 TGCCAGTGGTTTAAAGATTTGAGAGAACAGAGTATATGATGATGATGATGATGATGAT 1140  
DB 1028 TGCCAGTGGTTTAAAGATTTGAGAGAACAGAGTATATGATGATGATGATGATGATGAT 1087  
QY 1141 ATTGATCAACCAATTTATTTGAGAGCAAGAAATTTTCTTTGGAGCTTGTGTAAGAGC 1200  
DB 1088 ACTGATCAACCAATTTATTTGAGAGCAAGAAATTTTCTTTGGAGCTTGTGTAAGAGC 1147  
QY 1201 TGTGATATATATGAGGGAACCCAGCTGGGACATTCATTCGACAAATATGATCAAGGCTG 1260  
DB 1148 TGTGATATATATGAGGGAACCCAGCTGGGACATTCATTCGACAAATATGATCAAGGCTG 1207  
QY 1261 TAATATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320  
DB 1208 TAATATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1267  
QY 1321 TGTCAAAACAGATCAATATCTTATTTGATGATGATGATGATGATGATGATGATGATGAT 1380  
DB 1268 CTTCAAAACAGATCAATATCTTATTTGATGATGATGATGATGATGATGATGATGATGAT 1327  
QY 1381 TGGTCCAGAAATGAAGAGTTAATTTCTTCCAGAGATGCCATCAAGAGAACAGGCCCA 1440  
DB 1328 TGAACAGAGATGAAGAGTTAATTTCTTCCAGAGATGCCATCAAGAGAACAGGCCCA 1387  
QY 1441 AACCTTAATGTTAGTGAACCTTTTCCAGAGAAATTTCAAAGTTGGCTGCAAGTTT 1500  
DB 1388 AACCTTAATGTTAGTGAACCTTTTCCAGAGAAATTTCAAAGTTGGCTGCAAGTTT 1447  
QY 1501 AAAGTCAATATCTGTTTGTGCTGTTGAGCAAGTGGGTGAGCATGTAGAGATGTTCA 1560  
DB 1448 AAAGTCAATATCTGTTTGTGCTGTTGAGCAAGTGGGTGAGCATGTAGAGATGTTCA 1507  
QY 1561 GCGAGCCGTTCTCAAGTTGGCCAGTTCTCAAAAAGAGAAAGCTGTTGAATTTCTGCG 1620  
DB 1508 GCGAGCCGTTCTCAAGTTGGCCAGTTCTCAAAAAGAGAAAGCTGTTGAATTTCTGCG 1567  
QY 1621 AAACATAGGAGATGAAGAACTATGCTCTTGTGTAATTAAGAAAGAGAGATTTTAC 1680  
DB 1568 AAACATAGGAGATGAAGAACTATGCTCTTGTGTAATTAAGAAAGAGAGATTTTAC 1627  
QY 1681 TGCACCTTTCTTGTCAAGAAATATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 1740



Db 1628 TGCACTTTTCTTGTGTCAGAAAAATATCAAGATATTCATGTGATGGAGCA 1687  
Qy 1741 GAGGAGCGGAGAGCAAGCTCTGAGATTTTCGTTTGGAAAAAGTCCAGTTCTTTTCG 1800  
Db 1688 GAGGAGAGAGAGCAAGCTCTTGGAGATTTTCGCTGGAAAAAGTCCAGTTCTTTTCG 1747  
Qy 1801 TACTTCAATAGTCTGCCAGAGGGCTGATATGAAAAATGTCACATGTTATCAATTTGA 1860  
Db 1748 TACTTCAATAGTCTGCCAGAGGGCTGATATGAAAAATGTCACATGTTATCAATTTGA 1807  
Qy 1861 TCTTCTTCAATGATATGATATGTCATGCAATTTGGGCGTACTGGTCGTTGGGAA 1920  
Db 1808 CTTTCTTCAATGATATGATATGTCATGCAATTTGGGCGTACTGGTCGTTGGGAA 1867  
Qy 1921 TACTGAGAGCAATTTCTTTTGTATCTTGAATCGATTAACATTTAGCAGCGCTCT 1980  
Db 1868 TACTGAGAGCGCATTTCTTTTGTATCTGATCTGATTAATCATTTAGCAGCGCTCT 1927  
Qy 1981 AGTAAAGTATGACAGATGCTCAACAGAGATGTCCTGCAATGTTGAAAGAAATTCCTT 2040  
Db 1928 AGTAAAGTATGCTGACAGCGCTCAACAGAGATGTCCTGCAATGTTGAAAGATTCCTT 1987  
Qy 2041 TAGTACATACAT---TCTGAGCTTCAAGTGTGATACAGAA---GAAACGTTTTCATC 2094  
Db 1988 CAGTACTATGTGCTCTCCAGCTTTCAGTACAGCAAGAGGGGGGCTGTCTTTCATC 2047  
Qy 2095 AGTTGATACAGAA-----AGGCAAGAGCACTTTGAACACAGCTGGTTCCTTC 2145  
Db 2048 TGTGACACAGAGAGAAATTCACAGGCGCAAGCAACCTGAAATACAGCGGAGTTTCCTC 2107  
Qy 2146 TTCAAGAGCTCCCAATCCATGATATGATGATGATGATGATTAAGCA--AACATCTT 2203  
Db 2108 TTCAAGAGCTCCCAATCCATGATGATGATGATGATGATTAAGCAAGAACTATCTT 2167  
Qy 2204 CAACTGTGTGTTT 2218  
Db 2168 CAACTGTGTGTTT 2182

RESULT 3  
LOCUS CD557531 896 bp mRNA linear EST 11-JUN-2003  
DEFINITION AGENCOURT 14413281 NIH MGC 180 Homo sapiens CDNA clone  
IMAGE:30390465 5', mRNA sequence.  
ACCESSION CD557531  
VERSION CD557531.1 GI:31583599  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 896)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Daniela S. Gerhard, Ph.D.  
Office of Cancer Genomics / NIH  
National Cancer Institute  
Bldg. 31 Rm10A07 Bethesda, MD 20892  
Email: ggerds-remail.nih.gov  
Tissue Procurement: Dr. Michael Brownstein  
CDNA Library Preparation: Invitrogen Corp  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)  
DNA sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LINL at:  
http://image.lnl.gov  
Plate: NDAM465 row: j column: 10  
High quality sequence stop: 592.  
Location/Qualifiers  
1..896  
/organism="Homo sapiens"  
/mol\_type="mRNA"

/db xref="taxon:9606"  
/clone="IMAGE:30390465"  
/lab host="DH10B-Ton A (T1 and T5 phage resistances)"  
/clone\_id="NIH MGC 180"  
/note="Organ: Testis; Vector: PCMV-SPORE.1; Site 1: NotI;  
Site 2: EcoRV (destroyed); Library is oligo-dT primed and  
directionally cloned (EcoRV site is destroyed upon  
cloning). Average insert size 1.68 kb. Library was  
constructed by (Invitrogen). Note: this is a NIH\_MGC  
Library."

## ORIGIN

Query Match 34.7%; Score 772.2; DB 14; Length 896;  
Best Local Similarity 97.7%; Pred. No. 3.3e-193;  
Matches 794; Conservative 0; Mismatches 18; Indels 1; Gaps 1;  
Qy 919 CATTGCTAAAGCTGTTATTAAGCTTACTCTGTCGCAAAATACAGTATTCCTATCAT 978  
Db 1 CATTGCTAAAGCTGTTATTAAGCTTACTCTGTCGCAAAATACAGTATTCCTATCAT 60  
Qy 979 ACTTGCAAGAGAGATTGATGCTGTGCTCAAAACAGGCTGCGAAGACTGCGGCTTT 1038  
Db 61 ACTTGCAAGAGAGATTGATGCTGTGCTCAAAACAGGCTGCGAAGACTGCGGCTTT 120  
Qy 1039 TCTCTACCAATTTTGGCTCATGATGATGATGATGATGATGATGATGATGATGATGAT 1098  
Db 121 TCTCTACCAATTTTGGCTCATGATGATGATGATGATGATGATGATGATGATGATGAT 180  
Qy 1099 GTTGCAAGAGAGAGATTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1158  
Db 181 GTTGCAAGAGAGAGATTGATGATGATGATGATGATGATGATGATGATGATGATGAT 240  
Qy 1159 TTTGCAAGAGAGAGATTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1218  
Db 241 TTTGCAAGAGAGAGATTGATGATGATGATGATGATGATGATGATGATGATGATGAT 300  
Qy 1219 AACCCAGCTGGAGACATTCATTCGACAAATGTAACAGGCTGATATATATATGCTTAC 1278  
Db 301 AACCCAGCTGGAGACATTCATTCGACAAATGTAACAGGCTGATATATATATGCTTAC 360  
Qy 1279 TCTGCAAGAGAGAGATTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1338  
Db 361 TCTGCAAGAGAGAGATTGATGATGATGATGATGATGATGATGATGATGATGATGAT 420  
Qy 1339 CTTAGTTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1398  
Db 421 CTTAGTTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480  
Qy 1399 GTTAAATTTCTGCGAGAAATGCAATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1458  
Db 481 GTTAAATTTCTGCGAGAAATGCAATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540  
Qy 1459 AACTTTTCCAGAGAAATTCAGAGTTGGCTGCGAAGTTTAAAGTCAATATCTCTT 1518  
Db 541 AACTTTTCCAGAGAAATTCAGAGTTGGCTGCGAAGTTTAAAGTCAATATCTCTT 600  
Qy 1519 TGTGCTGTTGAGACAGTGGTGGAGCATGTAGATGTTGAGAGAGAGAGAGAGAGAGAG 1578  
Db 601 TGTGCTGTTGAGACAGTGGTGGAGCATGTAGATGTTGAGAGAGAGAGAGAGAGAGAG 660  
Qy 1579 TGGCCAGTTCTCAAAAAGAGAGAAAGCTGTTGAAATTCGCGAAACATAGGGAGAGAG 1638  
Db 661 TGGCCAGTTCTCAAAAAGAGAGAAAGCTGTTGAAATTCGCGAAACATAGGGAGAGAG 720  
Qy 1639 AACTAT-GGCTTGTGTAAGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1697  
Db 721 AACTATGGCTTTGTTGTAAGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780  
Qy 1698 AAGAAAAATATCAATCAAGTATTCATGCTG 1730  
Db 781 CAGAAAAATATTCATCAAGTATTCAGGGG 813

RESULT 4  
CD251828 850 bp mRNA linear EST 22-MAY-2003  
LOCUS  
DEFINITION  
AECNCOURT\_14206017 NIH\_MGC\_180 Homo sapiens cDNA clone  
IMAGE:30383198 5', mRNA sequence.  
ACCESSION  
CD251828  
VERSION  
CD251828.1 GI:31012294  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
NIH-MGC http://imgc.ncbi.nih.gov/  
1 (bases 1 to 850)  
NATIONAL INSTITUTES OF HEALTH, MAMMALIAN GENE COLLECTION (MGC)  
UNPUBLISHED (1999)  
CONTACT: DANIELA S. GERHARD, Ph.D.  
OFFICE OF CANCER GENOMICS  
NATIONAL CANCER INSTITUTE / NIH  
Bldg. 31 Rm10A07 Bethesda, MD 20892  
Email: cgaaps-remail.nih.gov  
TISSUE PROCUREMENT: Dr. Michael Brownstein  
CDNA LIBRARY PREPARATION: Invitrogen Corp  
CDNA LIBRARY ARRAYED BY: The I.M.A.G.E. Consortium (LINTL)  
DNA SEQUENCING BY: Agencourt Bioscience Corporation  
CLONE DISTRIBUTION: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LINTL at:  
http://image.llnl.gov  
Plate: NDAM46 row: k column: 15  
High quality sequence stop: 626.  
Location/Qualifiers  
1..850  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:30383198"  
/lab\_host="DH10B-Ton A ( T1 and T5 phage resistances) "  
/note="Organ: Testis; Vector: pCMV-SPORT6.1; Site 1: NotI;  
Site 2: EcoRV (destroyed); Library is oligo-dT primed and  
directionally cloned (EcoRV site is destroyed upon  
cloning). Average insert size 1.68 kb. Library was  
constructed by (Invitrogen). Note: this is a NIH\_MGC  
Library."  
ORIGIN  
Query Match 33.4%; Score 741.8; DB 14; Length 850;  
Best Local Similarity 98.1%; Pred. No. 3.7e-185;  
Matches 782; Conservative 0; Mismatches 12; Indels 3; Gaps 3;  
QY 10 CACATGCGGGGATGGAAGATTGGAGACGAAATCAACCTCATATGCTTCTCTATGTTCC 69  
DB 7 CCCCATGGAGAGATGGAGATTGGAGACGAAATCAACCTCATATGCTTCTCTATGTTCC 66  
QY 70 CATATTGAGAGAGATAGGATTTCTGGAGAAATGAGACATTTTAACAGAGATCCAGC 129  
DB 67 CATATTGAGAGAGATAGGATTTCTGGAGAAATGAGACATTTTAACAGAGATCCAGC 126  
QY 130 TTTCATCATGAAATGATGATGACCTTCTGAAAGATCATTTTCATGAAAAGTGAT 189  
DB 127 TTTCATCATGAAATGATGATGACCTTCTGAAAGATCATTTTCATGAAAAGTGAT 186  
QY 190 TGGCTTGGGCGGAAATTTTGGAAACGAGATGCTGGTGAATTAAGCGGATTAATAC 249  
DB 187 TGGCTTGGGCGGAAATTTTGGAAACGAGATGCTGGTGAATTAAGCGGATTAATAC 246  
QY 250 ATCCACATGAGTGTGTTTGGAGTTGGAAGAAGTTTGGAAAACAGAGTTTTCACAG 309  
DB 247 ATCCACATGAGTGTGTTTGGAGTTGGAAGAAGTTTGGAAAACAGAGTTTTCACAG 306  
QY 310 CAGGTTTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 369  
DB 307 CAGGTTTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 366

QY 370 TAAATCAACACGGAACAGAGGTTTCCAGAGAGCGGCTATGAGATGGAATATATTC 429  
DB 367 TAAATCAACACGGAACAGAGGTTTCCAGAGAGCGGCTATGAGATGGAATATATTC 426  
QY 430 AGAAGCTTCAGGCGCATATACAGAGAGGTTGGAAGAGTGTTCAGAGTTGCCGTGAGG 489  
DB 427 AGAAGCTTCAGGCGCATATACAGAGAGGTTGGAAGAGTGTTCAGAGTTGCCGTGAGG 486  
QY 490 ATTGGTCTAGAAAGTCCAAATTAATGACTTACACCCAGCAATGATATGACGCGCATCG 549  
DB 487 ATTGGTCTAGAAAGTCCAAATTAATGACTTACACCCAGCAATGATATGACGCGCATCG 546  
QY 550 TGGCTTTTGGTCTCTGGAAGACCAAGTATTAAGTGGACAGAGTAATGATATCTTCTCA 609  
DB 547 TGGCTTTTGGTCTCTGGAAGACCAAGTATTAAGTGGACAGAGTAATGATATCTTCTCA 606  
QY 610 AAGCAGAGTGCAGTGAAGTGAACGAGTGTTCACAAAGTTTAATGAAGAAGTAAT 669  
DB 607 AAGCAGAGTGCAGTGAAGTGAACGAGTGTTCACAAAGTTTAATGAAGAAGTAAT 666  
QY 670 AACAGGCTTCGGAAGAAGTTCTTGGAGTCAAGAAAGAGAGAGAGAAAGTGTGATAC 729  
DB 667 AACAGGCTTCGGAAGAAGTTCTTGGAGTCAAGAAAGAGAGAGAGAAAGTGTGATAC 726  
QY 730 TCAAGACCAAAAGTGAACCTTACA-TACCCCTCTCTCCACCTGAGAGATG-AGGACTCCATC 787  
DB 727 TCAAGACCAAAAGTGAACCTTACA-TACCCCTCTCTCCACCTGAGAGATGAGACTCCATC 786  
QY 788 -TTTGCACTTATTCAGA 803  
DB 787 TTTTGCACTTATTCAGA 803  
RESULT 5  
CD557866 933 bp mRNA linear EST 11-JUN-2003  
LOCUS  
DEFINITION  
AGENCOURT\_14423254 NIH\_MGC\_180 Homo sapiens cDNA clone  
IMAGE:30390020 5', mRNA sequence.  
ACCESSION  
CD557866  
VERSION  
CD557866.1 GI:31583934  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
NIH-MGC http://imgc.ncbi.nih.gov/  
1 (bases 1 to 933)  
NATIONAL INSTITUTES OF HEALTH, MAMMALIAN GENE COLLECTION (MGC)  
UNPUBLISHED (1999)  
CONTACT: DANIELA S. GERHARD, Ph.D.  
OFFICE OF CANCER GENOMICS  
NATIONAL CANCER INSTITUTE / NIH  
Bldg. 31 Rm10A07 Bethesda, MD 20892  
Email: cgaaps-remail.nih.gov  
TISSUE PROCUREMENT: Dr. Michael Brownstein  
CDNA LIBRARY PREPARATION: Invitrogen Corp  
CDNA LIBRARY ARRAYED BY: The I.M.A.G.E. Consortium (LINTL)  
DNA SEQUENCING BY: Agencourt Bioscience Corporation  
CLONE DISTRIBUTION: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LINTL at:  
http://image.llnl.gov  
Plate: NDAM46 row: g column: 21  
High quality sequence start: 19  
High quality sequence stop: 592.  
Location/Qualifiers  
1..933  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:30390020"  
/lab\_host="DH10B-Ton A ( T1 and T5 phage resistances) "  
/clone\_lib="NIH\_MGC\_180"

/note="Organ: Testis; Vector: pCMV-SPORT6.1; Site 1: NotI; Site 2: EcoRV (destroyed); Library is oligo-dt primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.68 kb. Library was constructed by (Invitrogen). Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 33.0%; Score 733.8; DB 14; Length 933;  
 Best Local Similarity 94.3%; Pred. No. 5.1e-183;  
 Matches 816; Conservative 0; Mismatches 42; Indels 7; Gaps 5;

979 ACTGCGAGACGAGATTGATGAGCTGTGCTCAACAGGCTGTGGAGAGACTCGGCTTT 1038  
 18 ATTCTCCGAGATGATGATGATGCTTGCCTTCAACGGCTCTGGAGAGCTCGGCTTT 77

1039 TCTCTCAACCAATTTGGCTCATATGATGATGATGATGATGATGATGATGATGATGAT 1098  
 78 TCTCTCAACCAATTTGGCTCATATGATGATGATGATGATGATGATGATGATGATGAT 137

1099 GTTGCAGAGACGAGAGTATTTGATGAGACCAACTCGAGATTTGGTCAACAGATTTTA 1158  
 138 GTTGCAGAGACGAGAGTATTTGATGAGACCAACTCGAGATTTGGTCAACAGATTTTA 197

1159 TTGGAAGCCAGAAAATTTCTTTTGGAGCTTGTGTAAGAGCTGTTGTAATATATATGAG 1218  
 198 TTGGAAGCCAGAAAATTTCTTTTGGAGCTTGTGTAAGAGCTGTTGTAATATATATGAG 257

1219 AACCCAGCTGGAGATTTCAATTCGACAAATATGATGATGATGATGATGATGATGATGAT 1278  
 258 AACCCAGCTGGAGATTTCAATTCGACAAATATGATGATGATGATGATGATGATGATGAT 317

1279 TCTGGAAGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1338  
 318 TCTGGAAGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 377

1339 CTGATGTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1398  
 378 CTGATGTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 437

1399 GTTAAATTTCTTCCCAAGAAATGATGATGATGATGATGATGATGATGATGATGATGAT 1458  
 438 GTTAAATTTCTTCCCAAGAAATGATGATGATGATGATGATGATGATGATGATGATGAT 497

1459 AACTTTTCCAGAGAAATTCAAAGTTGGCTGACAGATTTTAAAGTCAATATATCTGTT 1518  
 498 AACTTTTCCAGAGAAATTCAAAGTTGGCTGACAGATTTTAAAGTCAATATATCTGTT 557

1519 TGTGCTGTTGGAACAAGTGGTGAAGATGATGATGATGATGATGATGATGATGATGATGAT 1578  
 558 TGTGCTGTTGGAACAAGTGGTGAAGATGATGATGATGATGATGATGATGATGATGATGAT 617

1579 TGGCAGAGTTCTCAAAAAGAAAGCTGTGTAATCTGCGAAAATATAGGGGATG-AAA 1637  
 618 TGGCAGAGTTCTCAAAAAGAAAGCTGTGTAATCTGCGAAAATATAGGGGATG-AAA 677

1638 GAACATGATGCTTTGTTGTAACCTAAGAAAAGAGCAATTTTACTGCAACTTTTCTTTGTC 1697  
 678 GAACATGATGCTTTGTTGTAACCTAAGAAAAGAGCAATTTTACTGCAACTTTTCTTTGTC 737

1698 AAGAAAAAATATCAACTACAAAGATTCATGATGATGATGATGATGATGATGATGATGATGAT 1755  
 738 AAGAAAAAATATCAACTACAAAGATTCATGATGATGATGATGATGATGATGATGATGATGAT 797

1756 AGCTCTTGG-AGATTTTGGCTTTGAAAAGTCCCA-CTTCTTGTGCTACTT-CAGTAG 1811  
 798 AGCTCTTGGAAATTTTGGCTTTGAAAAGTCCCAAGTCTTGTGCTACTTCAATAGC 857

1812 CTGCGAGAGGGCTGATATGAAAA 1836  
 858 CGGCGAGAGGGCTGATATGAAAA 882

RESULT 6  
 CDS57160 885 bp mRNA linear EST 11-UN-2003  
 LOCUS  
 DEFINITION  
 AGENCOURT\_14413590 NIH\_MGC\_180 Homo sapiens cDNA clone  
 IMAGE:30387531 5', mRNA sequence.  
 CDS57160  
 ACCESSION  
 CDS57160  
 VERSION  
 CDS57160.1 GI:31583228  
 KEYWORDS  
 EST.  
 SOURCE  
 Homo sapiens (human)  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 1 (bases 1 to 885)  
 NIH-MGC http://mhc.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL  
 Unpublished (1999)  
 COMMENT  
 Contact: Daniela S. Gerhard, Ph.D.  
 Office of Cancer Genomics  
 National Cancer Institute / NIH  
 Bldg. 31 Rm10A07 Bethesda, MD 20892  
 Email: cgsabbs@mail.nih.gov  
 Tissue Procurement: Dr. Michael Brownstein  
 cDNA Library Preparation: Invitrogen Corp  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ULNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/ULNL at:  
 http://image.llnl.gov  
 Plate: NDM457 row: p column: 04  
 High quality sequence stop: 671.  
 Location/Qualifiers  
 1..885  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:30387531"  
 /lab\_host="DH10B-Ton A (T1 and T5 phage resistances)"  
 /clone\_11b="NIH\_MGC\_180"  
 /note="Organ: Testis; Vector: pCMV-SPORT6.1; Site 1: NotI;  
 Site 2: EcoRV (destroyed); Library is oligo-dt primed and  
 directionally cloned (EcoRV site is destroyed upon  
 cloning). Average insert size 1.68 kb. Library was  
 constructed by (Invitrogen). Note: this is a NIH\_MGC  
 Library."

## FEATURES

## source

ORIGIN  
 Query Match 32.8%; Score 729.6; DB 14; Length 885;  
 Best Local Similarity 99.0%; Pred. No. 6.5e-182;  
 Matches 765; Conservative 0; Mismatches 5; Indels 3; Gaps 3;

1223 GGAACATTCATTCGACAAATATGATGATGATGATGATGATGATGATGATGATGATGAT 1288  
 1 GGAACATTCATTCGACAAATATGATGATGATGATGATGATGATGATGATGATGATGAT 60

1289 CTGATGATATCATATGCGCAAGAAAGATTTGGTCCAAACAGATCAATATCTAGTTTGG 1348  
 61 CTGATGATATCATATGCGCAAGAAAGATTTGGTCCAAACAGATCAATATCTAGTTTGG 120

1349 GATGAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1408  
 121 GATGAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180

1409 TGGCCAGAAATGCCATCAAGAAAGACAGGCCCAACCTTATGTTGAGTCACTTTTCCA 1468  
 181 TGGCCAGAAATGCCATCAAGAAAGACAGGCCCAACCTTATGTTGAGTCACTTTTCCA 240

1469 GAGGAATTCAAAGTTGGCTCGAGAGTTTAAAGTCAATATCTGTTTGTGCTGTT 1528  
 241 GAGGAATTCAAAGTTGGCTCGAGAGTTTAAAGTCAATATCTGTTTGTGCTGTT 300

1529 GGAACATTCATTCGACAAATATGATGATGATGATGATGATGATGATGATGATGATGAT 1588



REFERENCE  
1 (bases 1 to 822)  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
Nih-MGC <http://mgc.nci.nih.gov/>.  
AUTHORS  
TITLE  
JOURNAL  
Unpublished (1999)  
National Institutes of Health, Mammalian Gene Collection (MGC)  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: Miklos Palokovics, M.D., Ph.D.  
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki  
Toshiyuki and Piero Carninci (RIKEN)  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
Plate: LMML1679 row: m column: 19  
High quality sequence stop: 714.  
Location/Qualifiers

FEATURES  
source  
1..822  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5269434"  
/lab\_host="DH10B"  
/clone\_lib="NIH MGC 97"  
/note="Organ: testis; Vector: pBluescript (modified  
pBluescript KS+); Site\_1: BamHI; Site\_2: SalI-XhoI  
(gagag); Oligo-dT primed using primer  
5'-TTTTTTTTTTTIVN-3', size-selected for average  
insert size 2.2 kb and normalized to R05. This is a  
primary library enriched for full-length clones and  
constructed using the Cap-trapper method (Carninci, in  
preparation). Library constructed by M. Brownstein  
(NIH/NHGRI, National Institutes of Health). Note: this is  
a NIH-MGC Library."

## ORIGIN

Query Match 29.3%; Score 652.2; DB 12; Length 822;  
Best Local Similarity 95.3%; Pred. No. 2,1e-161;  
Matches 737; Conservative 0; Mismatches 28; Indels 8; Gaps 6;

QY 1 ACTTGAAGTCCATGCGGAGTGAAGTGGGAAAGCAAGAAATCAACCTCATATGCTTC 60  
DB 51 ACTTGAAGTCCATGCGGAGTGAAGTGGGAAAGCAAGAAATCAACCTCATATGCTTC 110  
QY 61 CTATGTTCCCATATTTGAGAAAGATAGTATTTCTGAGAAATGAGACAAATTTTAAACAG 120  
DB 111 CTATGTTCCCATATTTGAGAAAGATAGTATTTCTGAGAAATGAGACAAATTTTAAACAG 170  
QY 121 GACTCCAGCTTCATCATGAAATGATGATGACCTTCTGAAAGAGATCATTTTATGAA 180  
DB 171 GACTCCAGCTTCATCATGAAATGATGATGACCTTCTGAAAGAGATCATTTTATGAA 230  
QY 181 AAGTGAATTTGCTCTGGGCGGAAATTTTGGAAAGAGATGCTGGTATGATTAAGCG 240  
DB 231 AAGTGAATTTGCTCTGGGCGGAAATTTTGGAAAGAGATGCTGGTATGATTAAGCG 290  
QY 241 AGATTAATATCCACAATGAGTGTGTTTGGAGTTGAAAGATTTTGAAGACAGAGTTT 300  
DB 291 AGATTAATATCCACAATGAGTGTGTTTGGAGTTGAAAGATTTTGAAGACAGAGTTT 350  
QY 301 TTCAAACGACGCTTTGAAGATGATGATGCTGCTGTTTCTGAGAGAGCTGTATATGA 360  
DB 351 TTCAAACGACGCTTTGAAGATGATGATGCTGCTGTTTCTGAGAGAGCTGTATATGA 410  
QY 361 CTGCGAAGATATCCACAACGAGACAGAGGCTTTTCCAAAGAGAGCGGCTATGAGATG 420  
DB 411 CTGCGAAGATATCCACAACGAGACAGAGGCTTTTCCAAAGAGAGCGGCTATGAGATG 470  
QY 421 AAATTAATTCAGAGCTTCAGGCGCATACAGAAAGGTGGA-AGAGGTAGTTTCCGAGTT 479  
DB 471 AAATTAATTCAGAGCTTCAGGCGCATACAGAAAGGTGGA-AGAGGTAGTTTCCGAGTT 530

QY 480 GCCGTGAG-AGGATTTGCTAGGAAGTCCAAATATGACTTAGACCCAGACAGATGATG 538  
DB 531 GCCGTGAGAGATTTGGTCTAGGAAGTCCAAATATGACTTAGACCCAGACAGATGATG 590  
QY 539 CAGCGCACTGGTGGCTTTTGTCTGAGAAACAGATTAATGAGGACAGTAATGCT 598  
DB 591 CAGCGCACTGGTGGCTTTTGTCTGAGAAACAGATTAATGAGGACAGTAATGCT 649  
QY 599 GATCTTCTCAAGCAGAGATGCGCAGTGGAAAGTGAACAGAGTGTATCAAAGTTAAT 658  
DB 650 GATCTTCTCAAGCAGAGATGCGCAGTGGAAAGTGAACAGAGTGTATCAAAGTTAAT 709  
QY 659 GAAGAGTATATACAGG---CTCTGAAAGAAATTTGGAAGTCAAG-AGGAGAGAGAG 714  
DB 710 GAAGAGTATATACAGGCTTTCTGGCAAGAAATTTGGAAGTCAAGAGAGAGAGAG 769  
QY 715 AGAAGTATGATATCTCAAG-ACCAAAAGTACCTACATACCCCTCTCTCA 766  
DB 770 AGAAGTATGATATCCCAAGGAAACCAAAAGTACCTACATACCCCTCTCTCA 822

RESULT 9  
CD557137 914 bp mRNA linear EST 11-JUN-2003  
LOCUS  
DEFINITION  
AGENCOURT 14413272 NIH MGC 180 Homo sapiens CDNA clone  
IMAGE:3030297 5', mRNA sequence.  
ACCESSION  
CD557137  
VERSION  
CD557137.1 GI:31583205  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Daniela S. Gerhard, Ph.D.  
Office of Cancer Genomics / NIH  
National Cancer Institute  
Bldg. 31 Rm10X07 Bethesda, MD 20892  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: Dr. Michael Brownstein  
CDNA Library Preparation: Invitrogen Corp  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
Plate: NDAM465 row: c column: 10  
High quality sequence start: 42  
High quality sequence stop: 582.  
Location/Qualifiers

## FEATURES

source  
1..914  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3030297"  
/lab\_host="DH10B-Ton A ( T1 and T5 phage resistances)"  
/clone\_lib="NIH MGC 180"  
/note="Organ: Testis; Vector: pCMV-SPORT6.1; Site\_1: NotI;  
Site\_2: EcoRV (destroyed); Library is oligo-dT primed and  
directionally cloned (EcoRV site is destroyed upon  
cloning). Average insert size 1.68 kb. Library was  
constructed by (Invitrogen). Note: this is a NIH-MGC  
Library."

## ORIGIN

Query Match 29.2%; Score 649; DB 14; Length 914;  
Best Local Similarity 96.6%; Pred. No. 1.5e-160;  
Matches 706; Conservative 0; Mismatches 20; Indels 5; Gaps 4;

QY 1080 CTGCCAGTGTGTTTAAAGATTGCAAGCAACGAGTGTATTTATGACCACTGAG 1139

Db 25 CGATTAATCGTTTAAAGAGTTCAGAGAACCAAGTGTATTTATTTAGACCAACCTCGAG 84  
Qy 1140 AATGTGCAACCAAGATTTA-TTTGAGAGCCAGAAAATTTCTTTGGGACCTTGTGAAGA 1198  
Db 85 AATTGTGCAACCAAGATTTA-TTTGAGAGCCAGAAAATTTCTTTGGGACCTTGTGAAGA 144  
Qy 1199 GCTGTGTATTAATATGAGGAGAACCCAGCTGGACATTCATTCAGCAAAATAGTACAGGC 1258  
Db 145 GCTGTGTATTAATATGAGGAGAACCCAGCTGGACATTCATTCAGCAAAATAGTACAGGC 204  
Qy 1259 TGTATATATATATATGAGTACTCTCTGGAGACATGATGATATCTATAGCAAAAGAAAGATT 1318  
Db 205 TGTATATATATATATGAGTACTCTCTGGAGACATGATGATATCTATAGCAAAAGAAAGATT 264  
Qy 1319 GGTCTCAACAGATCAAAATCTTATGTTTGGATGAAGCTGATGCAATGTTGAATAGGCT 1378  
Db 265 GGTCTCAACAGATCAAAATCTTATGTTTGGATGAAGCTGATGCAATGTTGAATAGGCT 324  
Qy 1379 TTTGGTCCAGAAATGAGAAATTAATTTCTTGCCCGAGAAATGCAATCAAGAAAGAGCCG 1438  
Db 325 TTTGGTCCAGAAATGAGAAATTAATTTCTTGCCCGAGAAATGCAATCAAGAAAGAGCCG 384  
Qy 1439 CAACACCTTATATGTCAGTGCACATTTTCCAGAGAAATTCAAAGGTTGCTGACAGATT 1498  
Db 385 CAACACCTTATATGTCAGTGCACATTTTCCAGAGAAATTCAAAGGTTGCTGACAGATT 444  
Qy 1499 TTTAATGCAATATATCTGTTTGTGCTTTGAGCAAGTGGTGGAGCATGTGAATGTT 1558  
Db 445 TTTAATGCAATATATCTGTTTGTGCTTTGAGCAAGTGGTGGAGCATGTGAATGTT 504  
Qy 1559 CACGACACGCTTCCAGAGTGGCCAGTTCTCAAAAGAGAAAGAGTGGTGAATTTCTG 1618  
Db 505 CACGACACGCTTCCAGAGTGGCCAGTTCTCAAAAGAGAAAGAGTGGTGAATTTCTG 564  
Qy 1619 CGAAACATAGGGGATGAAAGAACTATGCTTGTGTTGAATCTAGAAAAGACCAATTT 1678  
Db 565 CGAAACATAGGGGATGAAAGAACTATGCTTGTGTTGAATCTAGAAAAGACCAATTT 624  
Qy 1679 ACTGCAAA-CTTTCTTGTCAAGAAAATATCACTAAGATATCCAGTGTATC-663 1736  
Db 625 ATTGCAACCTTTCTTGTCAAGAAAATATCACTAAGATATCCAGTGTATC-663 684  
Qy 1737 AACGAGAGAGCGGAGCAAGCTCTTG--AGATTTTCTGCTTTGAAAGTGGCCAGTTCT 1794  
Db 685 AACGAGAGAGCGGAGCAAGCTCTTGAAATTTCTTGTGAAAGTGGCCAGTTCT 744  
Qy 1795 TGTGTGCTACTT 1805  
Db 745 CTGTGTGCTT 755

RESULT 10  
BQ429126 792 bp mRNA linear EST 24-MAY-2002  
LOCUS BQ429126  
DEFINITION AGENCOURT\_7901787 NIH\_MGC\_82 Homo sapiens cDNA clone IMAGE:6104663  
5', mRNA sequence.  
ACCESSION BQ429126  
VERSION BQ429126.1 GI:21168202  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 792)  
AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.  
TITLES National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapds-r@mail.nih.gov](mailto:cgapds-r@mail.nih.gov)  
Tissue Procurement: CLONTECH  
cDNA library Preparation: CLONTECH Laboratories, Inc.  
CDNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LICM2342 row: n column: 24  
High quality sequence stop: 496.  
Location/Qualifiers  
1..792  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6104663"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: testis; Vector: pDNR-LIB (Clontech); Site: 1:  
SfiI (ggccgctcgcc); Site 2: SfiI (ggccatctggcc); 5' and  
3' adaptors were used in cloning as follows: 5' adaptor  
sequence: 5'-CACGCCCATTAATGCG-3' and 3' adaptor sequence:  
5'-ATTCTAGAGCCGAGCGCCGACATG-dt(30)BN-3' (where B = A,  
C, or G and N = A, C, G, or T). Average insert size  
1.35 kb (range 0.9-4.0 kb). 14/15 colonies contained  
inserts by PCR. This library was enriched for full-length  
clones and was constructed by Clontech Laboratories (Palo  
Alto, CA)."

ORIGIN  
Query Match 28.9%; Score 643.2; DB 13; Length 792;  
Best Local Similarity 96.9%; Pred. No. 5e-159;  
Matches 677; Conservative 0; Mismatches 19; Indels 3; Gaps 2;  
Qy 1 ACTTAAGTACCAATGAGGAGATGAGATGAGAGCAAGATCAACCTCATATGCTTC 60  
Db 48 ACTTAAGTACCAATGAGGAGATGAGATGAGAGCAAGATCAACCTCATATGCTTC 107  
Qy 61 CTATGTTCCATATTTGAGAGAGATAGTATTTCTGAGAAATGAGCAATTTAAGC 120  
Db 108 CTATGTTCCATATTTGAGAGAGATAGTATTTCTGAGAAATGAGCAATTTAAGC 167  
Qy 121 GATTCAGCTTCATCATCAGAAATGAGATGAGACCTTCGAGAGATCATTTCAAGAA 180  
Db 168 GATTCAGCTTCATCATCAGAAATGAGATGAGACCTTCGAGAGATCATTTCAAGAA 227  
Qy 181 AAGTGAATTTCTTGGGGGGAATTTGGAAAAGAGATGCTGTGAGTAAATAGCG 240  
Db 228 AAGTGAATTTCTTGGGGGGAATTTGGAAAAGAGATGCTGTGAGTAAATAGCG 287  
Qy 241 AGATTAATACATCAAGATGAGGATTTTGGAGTTGGAAGAGTTTGGAAAAGAGATT 300  
Db 288 AGATTAATACATCAAGATGAGGATTTTGGAGTTGGAAGAGTTTGGAAAAGAGATT 347  
Qy 301 TTCAACAGAGGTTTGAAGATGATAGCTCTGTTTCTGAGAGAGCTAGTATAGA 360  
Db 348 TTCAACAGAGGTTTGAAGATGATAGCTCTGTTTCTGAGAGAGCTAGTATAGA 407  
Qy 361 CTGCGAAGATATCCAAACAGAGAGGTTTCCAAAGAGGCGGCTATTCAGATAG 420  
Db 408 CTGCGAAGATATCCAAACAGAGAGGTTTCCAAAGAGGCGGCTATTCAGATAG 467  
Qy 421 AAATTAATTCAGAGCTTCAGAGGCTATACAGAAAGAGTGAAGAGTATTCAGAGTTG 480  
Db 468 AAATTAATTCAGAGCTTCAGAGGCTATACAGAAAGAGTGAAGAGTATTCAGAGTTG 527  
Qy 481 CCGTGAAGATTTGCTCTAGAGAGTCCAAATTAATGATTAAGACCCAGAGAAATGATGCA 540  
Db 528 CCGTGAAGATTTGCTCTAGAGAGTCCAAATTAATGATTAAGACCCAGAGAAATGATGCA 587  
Qy 541 GCGCACTGGGCG--CGTTTGTGTTAGAGAGCAAGATTAATGAGTGAAGTGAAGTGAAG 599  
Db 588 GCGCACTGGGCGCTTTTGTGTTAGAGAGCCCTTAATTAAGTGAAGTGAAGTGAAG 647  
Qy 600 ATACTTCTCAAGCAGAGTGGAGTGAAGTGAAGCAGAGTGGTTACAAAGTTTAAATG 659  
Db 648 ATACTTCTCAAGCAGAGTGGAGTGAAGTGAAGCAGAGTGGTTACAAAGTGAATG 707

QY 660 AAGAGTAATACAGG--CTCTGGAAAGATTTCTGGAA 696  
 Db 708 AAGAGTAATACAGGCGCTCTGGAAAGATTTCTGGAA 746

RESULT 11  
 BG719449 747 bp mRNA linear EST 08-MAY-2001  
 LOCUS 602690172F1 NIH\_MGC\_97 Homo sapiens cDNA clone IMAGE:4822528 5',  
 DEFINITION mRNA sequence.  
 BG719449  
 VERSION BG719449.1 GI:13998636  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-remail.nih.gov  
 Tissue Procurement: Miklos Palikovits, M.D., Ph.D.  
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki  
 Toshiyuki and Piero Carninci (RIKEN)  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 cDNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.jnl.gov  
 Plate: LHAM0730 row: p column: 17  
 High quality sequence stop: 734.  
 Location/Qualifiers

FEATURES  
 source  
 1..747  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4822528"  
 /lab\_host="DH10B"  
 /clone\_id="NIH\_MGC\_97"  
 /note="Organ: testis; Vector: pBluescript (modified  
 pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI  
 (gtcgag); Oligo-dT primed using primer  
 5'-TTTTTTTTTTTTTTN-3', size-selected for average  
 insert size 2.2 kb and normalized to ROI 5. This is a  
 primary library enriched for full-length clones and  
 constructed using the Cap-trapper method (Carninci, in  
 preparation). Library constructed by M. Brownstein  
 (NIH/NHGRI, National Institutes of Health). Note: this is  
 a NIH\_MGC Library."

ORIGIN  
 Query Match 28.6%; Score 636.2; DB 12; Length 747;  
 Best Local Similarity 99.2%; Pred. No. 3.5e-157;  
 Matches 660; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

QY 1 ACTTGAGTCCATGGGGGATGAAGATTTGGGAAGCAACCTCATTTGCTTC 60  
 Db 84 ACTTGAGTCCATGGGGGATGAAGATTTGGGAAGCAACCTCATTTGCTTC 143

QY 61 CTATGTTCCATATTTGGAAGATAGTATTTGGAAGAAATGGACAAATTTTAA 120  
 Db 144 CTATGTTCCATATTTGGAAGATAGTATTTGGAAGAAATGGACAAATTTTAA 203

QY 121 GACTCCAGCTTCATCATCAAGAAATGATGAGACCTTTCCGAAGAGATCATTTCA 180  
 Db 204 GACTCCAGCTTCATCATCAAGAAATGATGAGACCTTTCCGAAGAGATCATTTCA 263

QY 181 AAGTGATTTGCTCTGGGCGGAATTTTGGAAACAGAGATGCTGTAGTGTATAAG 240  
 Db 264 AAGTGATTTGCTCTGGGCGGAATTTTGGAAACAGAGATGCTGTAGTGTATAAG 323

QY 241 AGATATATCATCCACATGSGTGTGTTTGGAGTTGAAAGATTTTGGAAACAGAG 300  
 Db 324 AGATATATCATCCACATGSGTGTGTTTGGAGTTGAAAGATTTTGGAAACAGAG 383

QY 301 TTCAAACAGCAGGTTTGAAGATGGATAGCTGTGTTTCTGGAGAGATCATGTA 360  
 Db 384 TTCAAACAGCAGGTTTGAAGATGGATAGCTGTGTTTCTGGAGAGATCATGTA 443

QY 361 CTGCGAGATATATCCACACCGAAACAGAGGTTTCCAAAGAGCGGCTATCGAG 420  
 Db 444 CTGCGAGATATATCCACACCGAAACAGAGGTTTCCAAAGAGCGGCTATCGAG 503

QY 421 AA-ATATTCAGAGCTTCAGGCGCATACAAAGAGGTGAAGGTATTTCCAGAG 479  
 Db 504 AACATTAATTCAGAGCTTCAGGCGCATACAAAGAGGTGAAGGTATTTCCAGAG 563

QY 480 GCCGTGAGAGATTTGGTCTAGAGAGTCCAAATTAATGACTAGACCACAGCAT 539  
 Db 564 GCCGTGAGAGATTTGGTCTAGAGAGTCCAAATTAATGACTAGACCACAGCAT 623

QY 540 AGCGCAGTGTGGCCTTTTGGTCTTGAAGACAGATTTAAGTGCACAGTAA 599  
 Db 624 AGCGCAGTGTGGCCTTTTGGTCTTGAAGACAGATTTAAGTGCACAGTAA 683

QY 600 ATACTTCTCAAGACGAAAGTGCAGTGAAGTGAACGAGTGTATCAAGATTA 659  
 Db 684 ATACTTCTCAAGACGAAAGTGCAGTGAAGTGAACGAGTGTATCAAGATTA 742

QY 660 AAGAA 664  
 Db 743 AAGAA 747

RESULT 12  
 BG718234 682 bp mRNA linear EST 08-MAY-2001  
 LOCUS 602696209F1 NIH\_MGC\_97 Homo sapiens cDNA clone IMAGE:4828298 5',  
 DEFINITION mRNA sequence.  
 BG718234  
 VERSION BG718234.1 GI:13997421  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-remail.nih.gov  
 Tissue Procurement: Miklos Palikovits, M.D., Ph.D.  
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki  
 Toshiyuki and Piero Carninci (RIKEN)  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 cDNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.jnl.gov  
 Plate: LHAM0746 row: a column: 03  
 High quality sequence stop: 682.  
 Location/Qualifiers

FEATURES  
 source  
 1..682  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4828298"  
 /lab\_host="DH10B"  
 /clone\_id="NIH\_MGC\_97"  
 /note="Organ: testis; Vector: pBluescript (modified  
 pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI  
 (gtcgag); Oligo-dT primed using primer



## ORIGIN

5'-TTTTTTTTTTTTTTN-3', size-selected for average insert size 2.2 kb and normalized to R05. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHRI, National Institutes of Health). Note: this is a NIH\_MGC Library."

Query Match 28.4%; Score 631.8; DB 12; Length 682;  
Best Local Similarity 99.7%; Pred. No. 4.9e-156;  
Matches 633; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ACTTGAAGTACCATGGGGATGAAATGGGAGGAAATCAACCTCATATGCTTC 60  
DB 48 ACTTGAAGTACCATGGGGATGAAATGGGAGGAAATCAACCTCATATGCTTC 107  
QY 61 CTATGTTCCCATATTTGAGAGAGATAGTATCTGAGAGAAATGAGACATTTTAACAG 120  
DB 108 CTATGTTCCCATATTTGAGAGAGATAGTATCTGAGAGAAATGAGACATTTTAACAG 167  
QY 121 GACTCCAGCTTCATCATCAAAATGATGATGACCTTCTCGAAGATCATTTTCATGAA 180  
DB 168 GACTCCAGCTTCATCATCAAAATGATGATGACCTTCTCGAAGATCATTTTCATGAA 227  
QY 181 AAGTGAATTTGCTGCGGCGAATTTTGGAAACAGAGATGCTGTAGTATATTAAGCG 240  
DB 228 AAGTGAATTTGCTGCGGCGAATTTTGGAAACAGAGATGCTGTAGTATATTAAGCG 287  
QY 241 AGATATATCATCAACATGGGTGCTTTGGAGTTGAAAGAGTTTGGAAACAGAGGTTT 300  
DB 288 AGATATATCATCAACATGGGTGCTTTGGAGTTGAAAGAGTTTGGAAACAGAGGTTT 347  
QY 301 TTCAAAACAGAGGTTTGAAGATGATAGCTCTGTTCTGAGAGATCTATGATATGA 360  
DB 348 TTCAAAACAGAGGTTTGAAGATGATAGCTCTGTTCTGAGAGATCTATGATATGA 407  
QY 361 CTGCGAAGATATCAACACAGAGAGGTTTTCGAAGAGCGGCTATCGAGATGG 420  
DB 408 CTGCGAAGATATCAACACAGAGAGGTTTTCGAAGAGCGGCTATCGAGATGG 467  
QY 421 AATTAATTCAGAGCTTGAAGGCTATCAAGAGAGTGGAAAGAGTATGTTCCGAGTTG 480  
DB 468 AATTAATTCAGAGCTTGAAGGCTATCAAGAGAGTGGAAAGAGTATGTTCCGAGTTG 527  
QY 481 CCGTGAAGATTTGCTCTAGAGATCAATATATGACTTACACCAAGATGATGCA 540  
DB 528 CCGTGAAGATTTGCTCTAGAGATCAATATATGACTTACACCAAGATGATGCA 587  
QY 541 GCGCACTGTGTGCTTTTGGTTCTAGAGACCAATTAAGTGGACAGGTAATGCTGA 600  
DB 588 GCGCACTGTGTGCTTTTGGTTCTAGAGACCAATTAAGTGGACAGGTAATGCTGA 647  
QY 601 TACTTCTCAAGCAGAGTGGCAGTGAAGTGAAC 635  
DB 648 TACTTCTCAAGCAGAGTGGCAGTGAAGTGAAC 682

RESULT 13  
BI560849  
LOCUS 603254018F1 NIH\_MGC\_97 Homo sapiens CDNA clone IMAGE:5296444 5',  
DEFINITION mRNA sequence.  
ACCESSION BI560849  
VERSION BI560849.1 GI:15448163  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 693)  
NIH-MGC http://mgc.nci.nih.gov/  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)

## JOURNAL COMMENT

Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cga@bbs-riemail.nih.gov  
Tissue Procurement: Mixlos Palkovits, M.D., Ph.D.  
cDNA Library Preparation: Michael J. Brownstein (NHRI), Shiroki  
Toshiyuki and Piero Carninci (RIKEN)  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.lnl.gov  
Plate: L14M11750 row: c column: 05  
High quality sequence stop: 691.

## FEATURES

## source

Location/Qualifiers  
1..693  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5296444"  
/lab\_host="DH10B"  
/clone\_lib="NIH\_MGC\_97"  
/note="Organ: testis; Vector: pBluescript (modified  
pBluescript KS+); Site\_1: BamHI; Site\_2: SalI-XhoI  
(gtagag); Oligo-dT primed using primer  
5'-TTTTTTTTTTTTTTN-3', size-selected for average  
insert size 2.2 kb and normalized to R05. This is a  
primary library enriched for full-length clones and  
constructed using the Cap-trapper method (Carninci, in  
preparation). Library constructed by M. Brownstein  
(NIH/NHRI, National Institutes of Health). Note: this is  
a NIH\_MGC Library."

## ORIGIN

Query Match 28.4%; Score 631.8; DB 12; Length 693;  
Best Local Similarity 99.7%; Pred. No. 5e-156; 2; Indels 0; Gaps 0;  
Matches 633; Conservative 0; Mismatches 2;

QY 1 ACTTGAAGTACCATGGGGATGAAATGGGAGGAAATCAACCTCATATGCTTC 60  
DB 59 ACTTGAAGTACCATGGGGATGAAATGGGAGGAAATCAACCTCATATGCTTC 118  
QY 61 CTATGTTCCCATATTTGAGAGAGATAGTATCTGAGAGAAATGAGACATTTTAACAG 120  
DB 119 CTATGTTCCCATATTTGAGAGAGATAGTATCTGAGAGAAATGAGACATTTTAACAG 178  
QY 121 GACTCCAGCTTCATCATCAAAATGATGATGACCTTCTCGAAGATCATTTTCATGAA 180  
DB 179 GACTCCAGCTTCATCATCAAAATGATGATGACCTTCTCGAAGATCATTTTCATGAA 238  
QY 181 AAGTGAATTTGCTGCGGCGAATTTTGGAAACAGAGATGCTGTAGTATATTAAGCG 240  
DB 239 AAGTGAATTTGCTGCGGCGAATTTTGGAAACAGAGATGCTGTAGTATATTAAGCG 298  
QY 241 AGATATATCATCAACATGGGTGCTTTGGAGTTGAAAGAGTTTGGAAACAGAGGTTT 300  
DB 299 AGATATATCATCAACATGGGTGCTTTGGAGTTGAAAGAGTTTGGAAACAGAGGTTT 358  
QY 301 TTCAAAACAGAGGTTTGAAGATGATAGCTCTGTTCTGAGAGATCTATGATATGA 360  
DB 359 TTCAAAACAGAGGTTTGAAGATGATAGCTCTGTTCTGAGAGATCTATGATATGA 418  
QY 361 CTGCGAAGATTTGCTCTAGAGATCAATATATGACTTACACCAAGATGATGCA 420  
DB 419 CTGCGAAGATTTGCTCTAGAGATCAATATATGACTTACACCAAGATGATGCA 478  
QY 421 AATTAATTCAGAGCTTGAAGGCTATCAAGAGAGTGGAAAGAGTATGTTCCGAGTTG 480  
DB 479 AATTAATTCAGAGCTTGAAGGCTATCAAGAGAGTGGAAAGAGTATGTTCCGAGTTG 538  
QY 481 CCGTGAAGATTTGCTCTAGAGATCAATATATGACTTACACCAAGATGATGCA 540  
DB 539 CCGTGAAGATTTGCTCTAGAGATCAATATATGACTTACACCAAGATGATGCA 598



QY 541 GCGACAGTGGCTTTTGGTTCTAGAACAGATTAAGTGGACAGATTAAGTGA 600  
 DB 599 GCGACAGTGGCTTTTGGTTCTAGAACAGATTAAGTGGACAGATTAAGTGA 658  
 QY 601 TACTTCTCAAGCAGAGAGTGGCAGTGGAGTGAAC 635  
 DB 659 TACTTCTCAAGCAGAGAGTGGCAGTGGAGTGAAC 693

RESULT 14  
 CD358073 868 bp mRNA linear EST 29-MAY-2003  
 LOCUS CD358073  
 DEFINITION AGENCOURT 14257996 NIH MGC 180 Homo sapiens CDNA clone  
 IMAGE:30386875 5', mRNA sequence.  
 ACCESSION CD358073 GI:31129484  
 VERSION CD358073.1 GI:31129484  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 868)  
 NIH-MGC http://mgs.nci.nih.gov/.  
 AUTHORS NIH-MGC  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Daniela S. Gerhard, Ph.D.  
 Office of Cancer Genomics  
 National Cancer Institute / NIH  
 Bldg. 31 Rm10A07 Bethesda, MD 20892  
 Email: cgabs-remail.nih.gov  
 Tissue Procurement: Dr. Michael Brownstein  
 CDNA Library Preparation: Invitrogen Corp  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.lnl.gov  
 Plate: NDAM456 row: d column: 20  
 High quality sequence stop: 578.

FEATURES  
 source  
 1..868  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:30386875"  
 /lab\_host="DH10B-Ton A ( T1 and T5 phage resistances)"  
 /clone\_lib="NIH MGC 180"  
 /note="Organ: Testis; Vector: pCW-SPORE1; Site 1: NotI;  
 Site 2: EcoRV (destroyed); Library is oligo-dT primed and  
 directionally cloned (EcoRV site is destroyed upon  
 cloning). Average insert size 1.68 Kb. Library was  
 constructed by (Invitrogen). Note: this is a NIH-MGC  
 Library."

ORIGIN  
 Query Match 28.4%; Score 631.2; DB 14; Length 868;  
 Best Local Similarity 96.6%; P: 0; Mismatches 19; Indels 5; Gaps 3;  
 Matches 677; Conservative 0; Mismatches 19; Indels 5; Gaps 3;

QY 1 ACTGAAGTCAACCTGGGGATGAAGATTGGAGCAGAAATCAACCTCATATGCTTC 60  
 DB 45 ACTGAAGTCAACCTGGGGATGAAGATTGGAGCAGAAATCAACCTCATATGCTTC 104  
 QY 61 CTATGTTCCATATTGGAAGATAGTATTCTGAGCAAAATGAGCAATTTTAACG 120  
 DB 105 CTATGTTCCATATTGGAAGATAGTATTCTGAGCAAAATGAGCAATTTTAACG 164  
 QY 121 GACTCAGCTTCATCATGAGAAATGATGAGACCTTCGAGAGATCATTTTCATGAA 180  
 DB 165 GACTCAGCTTCATCATGAGAAATGATGAGACCTTCGAGAGATCATTTTCATGAA 224  
 QY 181 AAGTGATTTGCTCTGGCGGAAATTTGAAAAGAGATGCTGTGATGTAAGCG 240

DB 225 AAGTGATTTGCTCTGGCGGAAATTTGAAAAGAGATGCTGTGATGTAAGCG 284  
 QY 241 AATATATCATTCACAAATGGGCTGTTTGGAGTGGAAAGAGTTTGGAAACAGAGTTT 300  
 DB 285 AATATATCATTCACAAATGGGCTGTTTGGAGTGGAAAGAGTTTGGAAACAGAGTTT 344  
 QY 301 TTCGAAGCAGAGTGTGAAGATGATAGCTCTGCTGTTTTCGAGAGATCTAGTAATGA 360  
 DB 345 TTCGAAGCAGAGTGTGAAGATGATAGCTCTGCTGTTTTCGAGAGATCTAGTAATGA 404  
 QY 361 CTGCGAAGATTAATCCACACGAGACAGAGGTTTTCGAGAGAGCGGCTATCGAGATGG 420  
 DB 405 CTGCGAAGATTAATCCACACGAGACAGAGGTTTTCGAGAGAGCGGCTATCGAGATGG 464  
 QY 421 AATATATTCAGAGGTTTTCGAGAGGTCATAGAGAGAGGTTTTCGAGAGGTTTTCGAGGTTG 480  
 DB 465 AATATATTCAGAGGTTTTCGAGAGGTCATAGAGAGAGGTTTTCGAGAGGTTTTCGAGGTTG 524  
 QY 481 CCGTGAAGATTTGGTCTAGAGAGTCCAAATATGACTTAGCCGACGAAATGATGCA 540  
 DB 525 CCGTGAAGATTTGGTCTAGAGAGTCCAAATATGACTTAGCCGACGAAATGATGCA 584  
 QY 541 GCGCACTGTGGCTTTTGGTTCTAGAGAGCCAGTATTAAGTGGCAGAGTAATAGTGA 600  
 DB 585 GCGCACTGTGGCTTTTGGTTCTAGAGAGCCAGTATTAAGTGGCAGAGTAATAGTGA 644  
 QY 601 TACTTCTCAAGCAGAGAGTGG-CAGTGGAGTGAACAGAGTGTATC--AAAGTTTAAA 657  
 DB 645 TACTTCTCAAGCAGAGAGTGGCAGAGTGGCAGAGTGTATC--AAAGTTTAAA 704  
 QY 658 TGAAGAGTATTAATACAG--CTCTGGAAGAAATTTTGGAA 696  
 DB 705 GGAAGAGTATTAATACAG--CTCTGGAAGAAATTTTGGAA 745

RESULT 15  
 BI462207 830 bp mRNA linear EST 21-AUG-2001  
 LOCUS BI462207  
 DEFINITION 60305520F1 NIH\_MGC\_97 Homo sapiens CDNA clone IMAGE:5271221 5',  
 mRNA sequence.  
 ACCESSION BI462207  
 VERSION BI462207.1 GI:15252863  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 830)  
 NIH-MGC http://mgs.nci.nih.gov/.  
 AUTHORS NIH-MGC  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgabs-remail.nih.gov  
 Tissue Procurement: Miklos Palokovits, M.D., Ph.D.  
 CDNA Library Preparation: Michael J. Brownstein (NHRRI), Shiraki  
 Toshiyuki and Piero Carninci (RIKEN)  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.lnl.gov  
 Plate: LLM11684 row: h column: 06  
 High quality sequence stop: 769.

FEATURES  
 source  
 1..830  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5271221"  
 /lab\_host="DH10B"  
 /clone\_lib="NIH MGC 97"  
 /note="Organ: testis; Vector: pBluescript (modified  
 pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI

(grecag); Oligo-dT primed using primer  
5'-TTTTTTTTTTTTTTTTVN-3', size-selected for average  
insert size 2.2 kb and normalized to ROP 5. This is a  
primary library enriched for full-length clones and  
constructed using the Cap-trapper method (Carninci, in  
preparation). Library constructed by M. Brownstein  
(NIMH/NHGRI, National Institutes of Health). Note: this is  
a NIH\_MGC Library."

## ORIGIN

```
Query Match      28.2%; Score 627; DB 12; Length 830;
Best Local Similarity 95.7%; Pred. No. 1e-154;
Matches 741; Conservative 0; Mismatches 21; Indels 12; Gaps 9;

QY 1 ACTTGAAGTACCATGGGGATGAGATTGGGAAGCAGAAATCACCCTCATATGCTTC 60
   |||
Db 59 ACTTGAAGCCACCATGGAGATGAGATTGGGAAGCAGAAATCACCCTCATATGCTTC 118
   |||

QY 61 CTATGTTCCCATATTATTGAGAAGATAGTATTCTGAGAAAAATGAGACAAATTTTACAG 120
   |||
Db 119 CTATGTTCCCATATTATTGAGAAGATAGTATTCTGAGAAAAATGAGACAAATTTTACAG 178
   |||

QY 121 GACTCCAGCTTCATCATCAGAAATGATGATGACCTCTCGAAGAGATTCATGAA 180
   |||
Db 179 GACTCCAGCTTCATCATCAGAAATGATGATGACCTCTCGAAGAGATTCATGAA 238
   |||

QY 181 AAGTGAATTTGCTCTGGGCGGAATTTGGAAAAGAGATGCTGTGAGTATATAAGCG 240
   |||
Db 239 AAGTGAATTTGCTCTGGGCGGAATTTGGAAAAGAGATGCTGTGAGTATATAAGCG 298
   |||

QY 241 AGATTAATACATCCACATGGGTGTTTGGAGTTGGAAAAGATTTGGAAAAGAGGTTT 300
   |||
Db 299 AGATTAATACATCCACATGGGTGTTTGGAGTTGGAAAAGATTTGGAAAAGAGGTTT 358
   |||

QY 301 TTCAACAGACAGGTTTGAGATGATGATGATGCTGTGTTCTGAGAGAGCTAGTATAGA 360
   |||
Db 359 TTCAACAGACAGGTTTGAGATGATGATGATGCTGTGTTCTGAGAGAGCTAGTATAGA 418
   |||

QY 361 CTGGGAAGATTAATCCAAACAGGAAGAGGTTTCCCAAGAGAGGCGCTATCGAGATGG 420
   |||
Db 419 CTGGGAAGATTAATCCAAACAGGAAGAGGTTTCCCAAGAGAGGCGCTATCGAGATGG 478
   |||

QY 421 AAATTAATTCAGAGCTTCAGGGCCATACAGAAAGGTGGAAGAGTAGTTCCGAGGTTG 480
   |||
Db 479 AAATTAATTCAGAGCTTCAGGGCCATACAGAAAGGTGGAAGAGTAGTTCCGAGGTTG 538
   |||

QY 481 CCGTGAAGCATTTGGTCTAGGAAGTCCAAATTAATGACTTAGACCCAGACCAATGTATGCA 540
   |||
Db 539 CCGTGAAGCATTTGGTCTAGGAAGTCCAAATTAATGACTTAGACCCAGACCAATGTATGCA 597
   |||

QY 541 GCGCAGCTGGTGCTTTTGGTCTTAGAAGACCAAGTATTAGT-GGCAAGGTAAATGGTG 599
   |||
Db 598 GCGCAGCTGGTGCTTTTGGTCTTAGAAGACCAAGTATTAGTGGGACAGGTAAATGGTG 656
   |||

QY 600 ATAATTCTCAAGCAGAAAGTGGCAG-TGGAAGTGAACG-AGGTGTTTCAAAAGGTTT-- 654
   |||
Db 657 ATAATTCTCAAGCAGAAAGTGGCAGTGGGAAGTGAACGAAGGTGTTTACCAAGGTTTCA 716
   |||

QY 655 AATGAAGAAGTATATAACAGGCTCTGGA-AGAACTCTGGAAGTCAGAA-GCAGAGGA 712
   |||
Db 717 ACTGAACGAAGTATATAACAGGCTCTGGAACAGCTTCTTGGGAAGTCAGAAAGCAGAGAGGA 776
   |||

QY 713 GGA-GAAAGTACTGATCTCAAGAGACCAAAAGTGACTTACATACCCCTCCTC 764
   |||
Db 777 GGAAGAAAGTACTGATCTCAAGAGACCAAAAGTGACTTATATACCCCTCCTC 830
   |||
```

Search completed: July 27, 2004, 10:49:42  
Job time : 5709 secs